

**THE VALIDATION OF NURSING MEASURES FOR PATIENTS WITH
UNPREDICTABLE OUTCOMES**

Hendrik Johannes Loubser

Student number: 516018

**A thesis submitted to the Faculty of Health Sciences, University of the
Witwatersrand, Johannesburg, and in fulfilment of the requirements for
the degree of Doctor of Philosophy.**

May 2012

I, Hendrik Johannes Loubser declare that this thesis is my own work. It is being submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.

A handwritten signature in black ink, consisting of a large, stylized initial 'H' followed by a horizontal line that ends in a small arrowhead pointing to the right.

May 2012

***Aan my familie: Antonio, Anli en Pierre en Lienke; Marieke en Albert en
Schalkie; Iza en Riekert en Annamie.***

Abstract

The sciences of restorative nursing are unknown in South Africa, leaving patients with restorative needs with rather unpredictable outcomes. This study investigated the validity of four prospective nursing scales to be used for patients requiring nursing where the focus is to improve their functionality. Such patients are usually found in sub- and non-acute nursing units and suffering with chronic debilitating diseases, mental illness or recovering from trauma. Typically they are in need of rehabilitation, palliative care, geriatric services or long-term care to restore or maintain their functional independence.

Inspired by the theories of nursing pioneers such as Florence Nightingale, the definitive nurse who was also an astute healthcare reform statistician, as well as Ida Jean Orlando, better known as the originator of the nursing process, the researcher, a general medical practitioner, has explored the intuitive knowledge of experienced nurses to document the links between their observations, interpretations and predictions of patient functioning. This information was used to develop four interrelated nursing scales to be used routinely by nurses to provide raw patient scores on patient functional changes. As nursing intuition was used to develop the measures, the working hypothesis was that the scales are considered valid. Therefore, the approach towards the study was deductive in nature, seeking the evidence to confirm this assumption.

As the purpose of the study was to offer nurses useful scales to provide validated empirical evidence of human functional status, the research question was how scientific evidence can be used to conclude that these four scales have indeed the integrity to deliver a measurement function to the nurses. The researcher's hypothesis of validating routine nursing measures is supported by two concepts: nursing utility and constructs validity. If nursing utility fail, construct validity is of no value to the nursing profession. With this in mind, the

study objectives were to first validate nursing utility using a qualitative design to collect descriptive data from nurses who have implemented the scales. Once positive findings were reported on the usefulness of the scales to the nursing profession, then construct validity was explored using the Rasch measurement model to qualitatively analyse the scale's raw data collected in various sub- and non-acute nursing facilities.

One scale was discarded, and three showed good to excellent results on both utility and construct validity. It has provided the restorative nursing sciences with a methodology to routinely collect patient-based empirical evidence for parametric analysis. In so doing, it delivered the missing link in Orlando's nursing process theory; it also confirmed Nightingale's theory that healthcare evidence provided routinely by nurse is the stepping stone for healthcare reform, provided it is useful, meaningful and valid. The ultimate beneficiaries of this new knowledge are patients who previously would have had unpredictable outcomes resulting in a poor prognosis.

COPYRIGHT

The copyright in this published edition is owned by the University of the Witwatersrand, Johannesburg. No part of this document may in any form or by any means (including without limitation: electronic, mechanical, photocopying, recording or otherwise) be used, copied, reproduced, amended, adapted, or transmitted. No express or any implied right or license to use, copy, reproduce, amend or adapt or transmit the whole or any part of this document is granted to any person or other legal entity.

The work embodied in this document is proprietary material of South African Database for Functional Medicine (Pty) Ltd and is covered by RSA patent application 2008/09086 and no part of the work may in any form or by any means (including without limitation: electronic, mechanical, photocopying, recording or otherwise) be used, copied, reproduced, amended, adapted, or transmitted, without the prior consent of South African Database for Functional Medicine (Pty) Ltd. No express or any implied right or license to use, copy, reproduce, amend or adapt or transmit the whole or any part of this document is granted to any person or other legal entity without the prior written consent of South African Database for Functional Medicine (Pty) Ltd, Box 2356, Houghton, 2041. Republic of South Africa.

Acknowledgements

When I wrote the last sentence in this thesis I have arrived at the end of a very long intellectual journey. For the feat of this adventure I am thankful to Professor Judy Bruce and Dr Daleen Casteleijn who made this journey a very personal experience. They never dictated, only questioned; never led, only guided and never prescribed, only suggested. For that I am very grateful.

I am also grateful to the many South African nurses who, during the past years, supported me on this road. I remain in awe of their science, their passion for patient wellbeing and their yearning for new care techniques. I am privileged to count them as my colleagues.

Thank you to my PhD travel-mate and friend, Marietha Nel, for co-piloting me through the numerous obstacles on this venture. Her editing of my manuscripts was particularly valuable.

Thank you to Elsie Venter who put me on the right track with the Rasch Measurement Model.

Finally I would like to thank my partner, Antonio Teixeira, for his presence, patience, perseverance and encouragement throughout this long journey. Without his technical support with the Rasch analyses, and help with the final editing and formatting of the manuscript, this study would not have been completed. To him I am eternally grateful.

TABLE OF CONTENTS

| | |
|--------------------------------------------------------------------------------|-----------|
| Chapter One: Introduction | 1 |
| 1.1 Introduction | 1 |
| 1.2 Conceptual framework of the study | 6 |
| 1.3 Background to the study | 6 |
| 1.4 The restorative nursing in context | 8 |
| 1.5 Research purpose | 13 |
| 1.6 Research problems | 13 |
| 1.7 Research objectives | 14 |
| 1.8 Scales, measures and numbers | 15 |
| 1.9 Significance of the study..... | 16 |
| 1.10 Definitions of key concepts..... | 17 |
| 1.11 Ethical considerations | 19 |
| 1.12 Summary..... | 20 |
| | |
| Chapter Two: Considerations | 21 |
| 2.1 Introduction | 21 |
| 2.2 Considering who qualifies to design and develop the nursing measures | 22 |
| 2.3 Considering the therapeutic contributions to the nursing measures | 22 |
| 2.4 Considering the context of restorative nursing practices | 23 |
| 2.5 Considering the interventions expected from restorative nursing | 25 |

| | | |
|---------------------------------------------------|-------------------------------------------------------------------------------|-----------|
| 2.6 | Considering the interface between acute nursing and restorative nursing | 27 |
| 2.7 | Considering nursing utility | 29 |
| 2.8 | Considering construct validity | 31 |
| 2.9 | Considering routine measures | 33 |
| 2.10 | Considering burden to apply, time and efficiency | 34 |
| 2.11 | Considering sustainability and focus on the end-user | 34 |
| 2.12 | Considering secondary use for healthcare reform | 35 |
| 2.13 | Summary | 38 |
| Chapter Three: Literature Review | | 39 |
| 3.1 | Introduction | 39 |
| 3.2 | Restorative nursing model | 39 |
| 3.3 | Nursing utility | 46 |
| 3.4 | Rasch measurement model (RMM) | 50 |
| | 3.4.1 Requirements for RMM analyses | 58 |
| | 3.4.2 Measurement expectations from the RMM | 59 |
| | 3.4.3 Fit statistic indices of the RMM | 60 |
| | 3.4.4 Category analyses guidelines of the RMM | 62 |
| 3.5 | Summary | 63 |
| Chapter Four: Methods | | 64 |
| 4.1 | Introduction | 64 |

| | |
|---------------------------------------------------------------|-----------|
| 4.2 Design for the entire study | 64 |
| 4.3 First study: Nursing utility | 65 |
| 4.3.1 Design for the nursing utility study | 66 |
| 4.3.2 Population, sample and setting | 68 |
| 4.3.3 Data collection | 69 |
| 4.3.4 Data quality | 70 |
| 4.3.5 Data analysis | 72 |
| 4.4 Second study: Construct validity | 74 |
| 4.4.1 Design for the construct validity study | 74 |
| 4.4.2 Population | 75 |
| 4.4.3 Settings | 75 |
| 4.4.4 Sampling | 76 |
| 4.4.5 Data collection | 77 |
| 4.4.6 Data quality | 77 |
| 4.4.7 Data analysis | 78 |
| 4.5 Rasch calibration | 81 |
| 4.5.1 Parameters used to report on category functioning | 84 |
| 4.5.2 Parameters used to report on item functioning | 86 |
| 4.6 Summary | 87 |
| Chapter Five: The ALPHA | 88 |
| 5.1 Introduction | 88 |
| 5.2 Development of the ALPHA | 89 |

| | | |
|----------------------------------------|---------------------------------------------------------|-----------|
| 5.3 | Conceptual framework of the ALPHA | 93 |
| 5.4 | Qualitative study: nursing utility | 94 |
| 5.5 | Quantitative study: construct validity | 95 |
| 5.5.1 | Sampling and data collection | 95 |
| 5.5.2 | Data collection for the Rasch analysis | 96 |
| 5.6 | Conclusion | 97 |
| Chapter Six: The BETA | | 98 |
| 6.1 | Introduction | 98 |
| 6.2 | Development of the BETA | 104 |
| 6.3 | Qualitative study: nursing utility | 107 |
| 6.3.1 | Sampling and data collection | 107 |
| 6.3.2 | Data analysis | 108 |
| 6.3.3 | Results from the professional nurses' focus group | 109 |
| 6.3.4 | Results from the caregivers' focus group | 112 |
| 6.4 | Quantitative study: construct validity | 115 |
| 6.4.1 | Sampling and data collection | 115 |
| 6.4.2 | Data preparation | 116 |
| 6.4.3 | Rasch calibration | 118 |
| 6.4.4 | Results on category functioning | 120 |
| 6.4.5 | Results on item functioning | 123 |
| 6.5 | Conclusion | 125 |

| | |
|---------------------------------------------------------------|------------|
| Chapter Seven: The GAMMA | 127 |
| 7.1 Introduction | 127 |
| 7.2 Development of the GAMMA | 131 |
| 7.3 Conceptual framework of the GAMMA | 133 |
| 7.4 Qualitative study: nursing utility | 134 |
| 7.4.1 Sampling and data collection | 134 |
| 7.4.2 Data analysis | 135 |
| 7.4.3 Results from the professional nurses' focus group | 136 |
| 7.5 Quantitative study: construct validity | 140 |
| 7.5.1 Sampling and data collection | 140 |
| 7.5.2 Data preparation | 142 |
| 7.5.3 Rasch calibration | 142 |
| 7.5.4 Results on category functioning | 144 |
| 7.5.5 Results on item functioning | 148 |
| 7.6 Conclusion | 149 |
| | |
| Chapter Eight: The DELTA | 152 |
| 8.1 Introduction | 152 |
| 8.2 Development of the DELTA | 155 |
| 8.3 Conceptual framework of the DELTA | 158 |
| 8.4 Qualitative study: nursing utility | 158 |
| 8.4.1 Sampling and data collection | 158 |

| | | |
|--------------------------------------------|---------------------------------------------------------|------------|
| 8.4.2 | Data analysis | 159 |
| 8.4.3 | Results from the professional nurses' focus group | 159 |
| 8.5 | Quantitative study: construct validity | 163 |
| 8.5.1 | Sampling and data collection | 163 |
| 8.5.2 | Data preparation | 164 |
| 8.5.3 | Rasch calibration..... | 165 |
| 8.5.4 | Results on category functioning | 168 |
| 8.5.5 | Results on item functioning | 172 |
| 8.6 | Conclusion | 173 |
| Chapter Nine: Conclusions | | 175 |
| 9.1 | Revisiting the origins and purpose of study | 175 |
| 9.2 | Conclusions of the study | 176 |
| 9.2.1 | Assumptions | 176 |
| 9.2.2 | Conceptual framework | 177 |
| 9.2.3 | Research objectives..... | 179 |
| 9.2.4 | Research problem | 179 |
| 9.2.5 | Research purpose..... | 180 |
| 9.3 | Reflections | 180 |
| 9.3.1 | Nursing intuition | 180 |
| 9.3.2 | Sustainability..... | 181 |
| 9.3.3 | Restorative nursing domain..... | 182 |
| 9.3.4 | Paucity of knowledge | 183 |

| | |
|------------------------------------------------|------------|
| 9.3.5 Routine data | 184 |
| 9.3.6 Patient outcomes | 185 |
| 9.4 Contribution to nursing knowledge..... | 185 |
| 9.5 Recommendations for nursing practice | 187 |
| 9.6 Recommendations for future research | 190 |
| 9.7 Final note | 191 |
| | |
| Bibliography | 192 |
| | |
| Annexure A..... | 210 |

List of Figures

- Figure 2.1:** The continuum of acute and restorative nursing episodes and settings. P24
- Figure 2.2:** Scope of functional changes in sub-acute restorative care. P26
- Figure 2.3:** Radar graph example of measure constructs. P32
- Figure 3.1:** The Category Probability Curve of the first DELTA item 1: Reality Loss. P62
- Figure 4.1:** Variable map of patient ability and item difficulty in the DELTA. P80
- Figure 4.2:** The Category Probability Curve of the DELTA scale's item 1: Reality Loss. P84
- Figure 5.1:** Radar graph representing the ALPHA scale structure. P91
- Figure 6.1:** Radar graph representing the BETA scale structure, based on the FIM platform. P107
- Figure 6.2:** Development of the four BETA sub-scales. P119
- Figure 6.3:** Variable map of patient ability and item difficulty in the BETA (Subscale: Self-care). P120
- Figure 6.4:** Examples of the BETA category probability curves before and after collapsing. P122
- Figure 7.1:** Radar graph representing the GAMMA scale structure. P134
- Figure 7.2:** Variable map of patient ability and item difficulty in the GAMMA scale. P144
- Figure 7.3:** Examples of the GAMMA category probability curves before and after collapsing. P147
- Figure 8.1:** Polar graph representing the Delta scale design. P157
- Figure 8.2:** The Category Probability Curve of Item 1: Reality Loss. P166
- Figure 8.3:** Variable map of patient ability and item difficulty in the DELTA. P168
- Figure 9.1:** BETA scale with admission and predicted outcome score P188

List of Tables

Tables

Table 4.1: Overall research design of the study. P65

Table 4.2: Structured questions for the focus groups. P70

Table 4.3: Parameters used in the calibration of category functioning. P85

Table 4.4: Parameters used in the calibration of item functioning. P86

Table 5.1: Outline of the ranking of the ALPHA categories. P92

Table 6.1: The new structure of the BETA categories that required collapsing. P121

Table 6.2: Results on the BETA item functioning. P124

Table 7.1: Results on the GAMMA category functioning. P146

Table 7.2: Results on the GAMMA item functioning. P149

Table 8.1: Description of DELTA items. P156

Table 8.2: Results on the DELTA category functioning. P169

Table 8.3: Results on the DELTA item functioning. P172

CHAPTER ONE: INTRODUCTION

"If the function of a hospital is to kill the sick, statistical comparisons of this nature (mortality) would be admissible. As, however, its proper function is to restore the sick to heal as speedily as possible, the elements which really give information as to whether this is done or not, are those which shows the proportion of sick restored to health and the average time which has been required for this object."

Florence Nightingale in her "Notes on Hospitals" 1863

1.1 Introduction

A process is a series of actions or operations leading towards a measurable goal. The nursing process is a nursing model or philosophy, which aims to provide patients with appropriate care, cure and comfort. It involves a series of five distinct actions or phases (Gillies 1982), and it is based on a theory developed by Ida Jean Orlando (Orlando 1961). She observed nurses in action in the late 1950's and reported that she saw "good" nursing and "bad" nursing. From her observations she formulated a framework for "good" nursing and it became known as the nursing process which, over time, became the gold standard of the nursing sciences. Orlando's teachings included three fundamentals:

- *The patient must be the central character of the nursing process*
- *The nursing care needs must be directed at improving outcomes for the patient; it is not about nursing goals*
- *The nursing process is an essential part of the nursing care plan*

Today the nursing process involves five distinct sequential phases starting with a nursing assessment of the patient's needs, then assimilating the information to form a nursing diagnosis, conceptualising a nursing plan to address the need,

implementing the appropriate nursing interventions or actions, and finally evaluating the patient outcome as a result of the nursing intervention. Although the nursing process is a sound model, it has limitations when applied to patients requiring restorative nursing. Under these circumstances, Orlando's assessment and evaluation phases have no patient-evidence based measurements to underpin its precision, and as a result the nursing quality indicators have not become patients-outcomes based as foreseen by Orlando ([Orlando & Dugan 1989](#)), but rather nursing practice-outcomes based. This lack of measurement invalidates the restorative nursing process as it has no means to empirically quantify the assessment and evaluation phases of the nursing process. The motive of this study is to rectify this situation.

Orlando's generic nursing process can be applied in both the curative and the restorative domains of the nursing sciences. In the curative domain of nursing the focus may be on immobilisation to optimise the bodily organs and systems to cure themselves and the parameters of success are usually collaborated empirically with vital signs and laboratory reports. In the restorative domain the nursing focus changes towards active mobilisation and functional gain with the objective to optimise personal independence. However, an extensive literature review ([PubMed](#), [Ebsco: CINAHL](#) and [JSTOR](#)) yielded no nursing scores (measures, instruments or tools) to empirically quantifying human function and to underpin the restorative nursing process. A similar finding was confirmed by [Smith and MacVicar \(1999 p 394\)](#), described as follows:

"Nurses are well prepared to respond to the challenge of providing appropriate restorative training and rehabilitation within a managed care environment. Although interventions aimed at improving mobility and increasing physical activity is within the scope of nursing practice, many nursing interventions lack the scientific underpinnings or empirical evidence necessary to document the desired outcomes."

Florence Nightingale (1820 -1910), the iconic English nursing pioneer, also hailed as the "*lady of the lamp*", is revered by nurses all over the world as the

founder of their profession. Based on her experiences in 1854 as a lay carer in the military hospitals during the Crimean war, Nightingale felt a calling to train ordinary women to become skilled nurses caring for the ill and wounded. In 1860 she opened the first ever official nurses' training program in London, the Nightingale School for Nurses. Nightingale is not only remembered for being a nursing pioneer, she was also an ardent collector of hospital-based outcomes mortality data, a prolific writer and she had exceptional skills to visually present her research information in statistical graphs novelty for that period. She wrote: *"I want everyone to understand - no hiding behind the supposed incomprehensibility of statistics. The figures must be as clear as a picture; they must tell a story as clearly as does a picture of the Crucifixion"* (Open University 1955). These qualities equally contributed to the Nightingale legacy in healthcare sciences, but relevant to this study is her visionary insistence on pursuing all methods to collect accurate evidence-based data for her statistical reporting and on which to base her nursing training programs. Nightingale, as a nursing tutor, considered clinical evidence as the most important tool nurses have for decision-making. She wrote: *"(Statistical) evidence, which we (nurses) have, means to strengthen for or against a proposition, is our proper means for attaining truth"* (Nightingale 1860 p 58).

Nightingale, the nurse, saw the wrong in the hospital systems of that period, and she knew what reforms it would require to improve the mortality statistics. She lamented: *"There was a growing conviction that in all hospitals there was a great and unnecessary waste of life. In attempting to arrive at the truth, I applied everywhere for information, but in scarcely an instance have I been able to obtain hospital records fit for any purposes of comparison"* (Nightingale 1863 p 175). However, she persevered and based on mortality statistics she set out on a path to convince those in power of the necessity of her proposed reforms for hospital and nursing services. Throughout all her writings she argued strongly that only by collecting and analysing pertinent data was it possible to determine the extent to which hospitals were effective in serving those who relied on their help. The statistical knowledge she gained and disseminated became the basis for her "effective hospital" campaign and the improvement of nursing services.

As relentless as her efforts were, working for reform in hospital services, as meticulous did her research become. According to her writings she became an ardent collector of accurate patient data, and when these were unavailable or inadequate she pressed for data collection. Her quest for data collection became a strong theme in her *Notes on Hospitals* (Nightingale 1863) and in her *Introductory Notes on Lying-In Institutions* (Nightingale 1871). With appropriate evidence-based data, her statistical analytic skills and her ability to enlighten authorities with graphical presentations, she perfected the process of healthcare reform. She always quoted figures to back up her case for reform: *“I collected my figures with a purpose in mind, with the idea that they could be used to argue for change. Of what use are statistics if we do not know what to make of them? What we wanted at that time was not so much an accumulation of facts, as to teach the men who are to govern the country the use of statistical facts”* (Cook 1913 p 396).

As Nightingale, the statistical analyst, explored beyond the use of mortality data to understand hospital outcomes, she discovered new probabilities and the relevance of her vision to her study became more apparent. She became aware of the variability in medical practices and the consequences this had on patient outcomes. She also became acutely aware of the need to adjust risk when comparing the outcomes of different groups of patients in different hospitals. Her interest shifted from hospital-based outcomes using mortality data to patient-based outcomes using performance data. She wrote: *“In comparing the deaths of one hospital with those of another, any statistics are justly considered absolutely valueless which do not give the ages, the sexes and the diseases of all the cases. There can be no comparison between old men with dropsies and young women with consumptions”* (Nightingale 1859 p 97). To get a better understanding of these patient variables she put forward a strong case for *“some uniform system of reporting”* on patient outcomes in hospitals. She pleaded her case as follows: *“The proportion of recoveries, the proportion of deaths, and the average time in hospital, must all be taken into account in discussions of this nature, as well as the character of the cases and the*

proportion of different ages among the sick. For me, this experience emphasised the great importance of correct hospital statistics as an essential element in hospital administration” (Nightingale 1863 p 5).

She religiously believed that patient-evidence-based statistical reporting, and particularly the use of diagrams to illustrate such evidence-based outcomes is a method to get a better understanding of God’s interventions in healthcare outcomes: *“The true foundation of theology is to ascertain the character of God. It is by the aid of such diagrams in particular, and Statistics in general that the law in the social spheres can be ascertained and codified, and certain aspects of the character of God thereby revealed. The study of (patient-outcomes) statistics is thus a religious service” (David 1963 p 103).*

Today these same theories she preached and practiced in the 19th century still hold as the same basic premise to sustain evidence-based approaches in modern medicine, in public health and in nursing. The principles she considered vital in healthcare reform in the 19th century are the same principles rediscovered in the modern healthcare literature.

“The principles remain the same, no matter how many zero’s you add to the problem” (Joffe 2009).

Nightingale was not only the first nurse history recognises, she was also the first researcher in history promoting exploring and reporting on patient evidence-based outcomes. Thus, from her work it is evident that she was also the first clinician in history to set systems in place to harness the unpredictability of patients’ outcomes. This study places the work of Florence Nightingale as a patient-evidence-based researcher on record and intends to contribute to her vision by validated nursing measures for patients with unpredictable outcomes.

1.2 Conceptual framework of the study

The conceptual framework that formed the cohesive thread throughout this study is based on the researcher's hypotheses that improvement of patients with unpredictable outcomes (e.g. those requiring restorative nursing) is directly linked to the patients' functional gains over time. Therefore, if the patient's functional status can be longitudinally quantified over nursing days, the patient improvement between admission and discharge can be numerically calculated. Moreover, if similar groups of patients' improvement patterns (e.g. strokes or spinal injuries), could be statistically established from pooled data certainly such summed calculations are a reflection of the effectiveness and efficiencies of restorative nursing services rendered. Furthermore, with patient-evidence based data available to provide empirical evidence of effective and efficient nursing performance, one can assume that the quality of restorative nursing can be directly related to patient outcomes. Lastly, one can postulate that the nursing audits for quality nursing may in future also consider the empirical evidence of patient outcomes.

1.3 Background to the study

With human function already being the universal construct of measure in the therapeutic sciences (Fawcett 2007), the researcher facilitated a series of interviews and investigative processes with nursing practitioners to design four nursing measures on similar principles but based on the existing restorative nursing processes. Although these four measures have different domains, interconnected, they represent a holistic picture of human functioning which is easily observable by the nurse caring for patients. These observations are typically seen in rehabilitation, convalescent, geriatric and mental healthcare practices. The four nursing measures are named ALPHA, BETA, GAMMA and DELTA. They are based on the World Health Organisation's International Classification of Functioning, Disability and Health ([ICF 2001](#)).

- The ALPHA scores the functioning of bodily structures and systems. It aims to represent the “*body functions and structures*” component in the ICF structure (ICF 2001). It is considered to be used by the nurse in the general ward who has to establish if the patient is ready to be discharged to the sub-acute section for restorative nursing care. The ALPHA is designed to answer the question: What is the empirical evidence that the patient’s organs and systems have stabilised sufficiently to endure active restorative nursing towards functional independence? The Alpha has twelve items and seven categories to each item (see Annexure A). Chapter Five of this study is devoted to the development and validation of the ALPHA.
- The BETA scores the Activities of Daily Living (ADLs) and it represents the activities component in the ICF structure as denoted by the prefixes d4 “*communication*”, d5 “*mobility*” and d6 “*self-care*” (ICF 2001). It is designed to be used routinely by the primary caregivers (e.g. nursing assistants or care givers) to score the patient as the ADLs are performed in all nursing facilities where ADLs are performed by caregivers, such as rehabilitation, convalescent care, geriatric frail care, institutional care, etc. The BETA is designed to answer the question: What is the empirical evidence of the person’s ability to perform his/her ADLs? The BETA has eighteen items and seven categories to each item (see Annexure A). Chapter Six of this study is devoted to the development and validation of the BETA.
- The GAMMA scores the Instrumental Activities of Daily Living (IADL’s). It also represents the activities component in the ICF, but at a higher functional level such as denoted by the ICF prefixes of d2 “*general tasks and demands*” and d6 “*domestic life*” (ICF 2001). In essence it refers to the independent living abilities of a patient. It is designed to be used routinely by nurses visiting patients in their homes, such as home based nursing, retirement village nursing, etc. The GAMMA is designed to answer the question: What is the empirical evidence of the person’s ability live independently? The GAMMA has eight items and seven

categories to each item (see Annexure A). Chapter SEVEN of this study is devoted to the development and validation of the GAMMA.

- The DELTA scores the executive functioning of the patient. It represents the highest activity level in the ICF as denoted by d7 “*interpersonal interaction and relationships*”, d8 “*major life areas*” and d9 “*community, social and civic life*” (ICF 2001). The DELTA is most helpful with the nursing care of acute mental healthcare illness. A low DELTA score indicates a severely ill patient. The DELTA is designed to be used routinely by mental healthcare nurses in an acute mental healthcare nursing facility. The DELTA is designed to answer the question: What is the empirical evidence of how severely ill the person is with acute mental health illness? The DELTA has five items and seven categories to each item (see Annexure A). Chapter EIGHT of this study is devoted to the development and validation of the GAMMA.

With all four of the nursing measures designed and developed, the next logical phase is the implementation. But these scales cannot be presented to the nursing profession if not validated. The researcher’s working hypothesis is that the scales will be found useful by the nursing profession and the scales will perform as fundamental measures providing accurate data to make inferences regarding patient improvement and quality nursing services. Thus, both nursing utility and the construct validity of each scale require investigation, analysis and reporting. If a scale indicates poor nursing utility, it will be rejected by the nursing profession as a routine measure, and as a result no longitudinal data will be collected, irrespective of high accuracy levels attained. Therefore, with nursing utility validated, an investigation into the scales’ construct validity becomes relevant. Only if both these validated properties can be achieved, will the probability exist that nurses will collect routine longitudinal patient-based outcomes data for current practice and future research. This study is designed to investigate both of these phenomena, viz., nursing utility and construct validity, before the measures can be introduced to the nursing profession as validated nursing measures.

1.4 Restorative nursing in context

Nurses in acute care settings correctly use the curative nursing process. This process is well structured with numerous quantitative measures to underpin and guide patient improvement, such as the vital biomedical data, laboratory and other special investigations. The curative nursing process is therefore quantifiable using fundamental measures providing measurement data to statistically analyse patient outcomes. However, once these measures indicate that the patients' organs and systems function optimally, the patients' are declared to be stable, meaning the acute phase is over and the curative nursing process has come to an end. If there is a residual motor, cognitive or behavioural functional loss as a consequence of the acute event, the patient is referred to sub-acute settings for rehabilitation, convalescent, palliative or psychiatric nursing services. At this point of triage, the nursing process and service change dramatically from the curative nursing model where the patient is actively immobilised to the restorative nursing model where the patient is actively mobilised according to a structured nursing care-plan. But, without patient-based measures to underpin and guide the restorative nursing process and patient outcomes, the restorative nursing process is at a loss to make accurate assessments, conclude a nursing diagnosis, develop a nursing care plan and evaluate the outcome. This phenomenon was identified when the relatively new nursing practice of sub-acute care was introduced in South Africa. It seems that this lack of patient based functional outcome measures also exists in standard British nursing practice ([Le May & Williams 2006](#)).

The main objective of care in sub-acute settings is to restore patients' collateral motor, cognitive and behavioural functional loss after an acute or chronic episode. However, the inability of the nursing process to replace the biomedical parameters with functional parameters to quantify and record a patient's functionalities became evident, and therefore the patient in the sub-acute and non-acute settings became known to the managed care industry as the patient with an unpredictable outcome.

The same lack of quantifiable measures for patient outcomes was observed by the researcher in home-based care and long-term care facilities where the nursing focus should be on facilitating functional gains for their patients. As nurses in these facilities also have no method or process of measuring their patients' functional status, they end-up with nondescript assessments leading to inconclusive diagnoses. With this resulting uncertainty, hesitancy and doubt, these nurses inherently revert to the known classical curative model of nursing by immobilising the patient, resulting in very high levels of dependency and frailty in high risk patients.

The inability to measure and manage patients' functional status appropriately has wider implications. Without a method to measure, communicate and report on patient function, the nursing profession is effectively excluded from multidisciplinary team discussions to assess the patient with unpredictable outcomes, to set meaningful and measurable goals for the patient and to analyse quality of nursing care, based on patient outcomes. Whilst the therapeutic team members seek to increase independence the traditional nurses, not previously exposed to restorative nursing principles, inherently revert to the known concepts of rendering total immobilised care and using vital and bio-statistics as outcomes measures.

The researcher has confirmed the absence of nursing measures on human function by means of a literature survey and personal interviews with expert nursing managers, academics and educationists such as Dr S Anderson ([Anderson 2010](#)), Prof MC Herbst ([Herbst 2010](#)), Prof S Human ([Human 2010](#)), Prof M Clarke ([Clarke 2010](#)) and Dr T Heyns ([Heyns 2010](#)). Currently the only functional measures known are the "Apgar" for neonates, the "Waterlow" for pressure sores, the "Morse Fall Scale" for patients at risk of falling and the "Glasgow Coma Scale" to record the conscious state of a patient. All of these nursing measures, except for the Apgar, are impairment based and not useful as generic measures for patients with unpredictable outcomes.

At this point, it is necessary to mention that the private hospital groups in South Africa have each developed their own in-house acuity measures used by their nursing staff. However, these scales are not patient-based outcomes scales, but are rather patient-based utilisation tools, without any validation properties, used to record the resources utilised by the patient on a daily basis in an attempt to quantify the operational costs to be charged by the facility to the medical scheme. These utilisation tools do not collect data on patient function and therefore cannot explain patient progress or analyse the clinical performance of nursing or multidisciplinary practices.

In an attempt to overcome the problem of a lack of patient-based measures, adapted occupational therapy scores (e.g. the FIM and the ROMS) have been implemented by some South African sub-acute care facilities. These therapeutic based scores have little clinical utility for the nursing profession as they are neither validated as a nursing measure nor embedded in the nursing records and processes. These scores are also not rooted in the nursing sciences and do not offer a clear benefit to the nursing profession. Furthermore, these scores have not been successful in replacing the current descriptive text assessments in the daily nursing report. Nurses therefore view these measurement tools as an additional administrative burden.

Ironically, from a pragmatic point of view, the nursing profession is better placed than the therapeutic professions to observe and record longitudinal changes in patient functioning as it happens from day-to-day. The nurse being continuously present as the primary caregiver is thus able to provide actual scores, whilst the therapists only have limited windows of access, mostly in simulated environments. For this reason, nurses refer to therapist observations as potential scores. This conflict between actual and potential scores may create malfunctioning multidisciplinary teams, where the nurses' attempts to provide scores are overseen by therapists. Therapists' view the lower actual scores provided by nurses as inaccurate and reason that it is due to the lack of nursing exposure to the therapeutic processes and an inherent inability of nurses to implement the therapeutic based measures correctly. Ultimately, patient outcomes must be

based on actual performance data and nurses should provide this data because they observe the patient over 24 hours. However, this potential contribution of valuable data by multidisciplinary team in the nursing profession is currently not possible since the measures are in the domain of, and controlled by, the therapeutic professionals. It does not belong to the nursing professionals.

However, if the scales nursing utility is of such a nature that it will entice the restorative nurses to record and chart patient functional gains routinely, nurses may provide the evidence of the sub- and non-acute healthcare outcomes. The pooled patient data may become a rich research platform with interval level data allowing metric statistics to explore the variables that influence patient improvement patterns (Bond & Fox 2007). The nursing patient records may become the source to investigate and produce answers to long outstanding research questions, such as:

- Does early onset rehabilitation improve patient outcomes?
- What is the effect of age, ethnicity and gender on impairment outcome?
- What functional benchmarks dictate a patient's triage to alternative levels of care?
- Does more than one functional related grouping exist within one impairment group and if so, what can we learn from it?
- What is the relationship between a nursing technique and patient outcome?
- Can we statistically prove over time that similar patients with similar impairments given similar treatments will have similar outcomes?
- If these relationships between variables can be proven, what would the effect be if the treatment modalities are manipulated?
- Is it possible to statistically establish that the "unpredictable" nature of impairments is a misnomer?

In future, with statistical analysis of validated data provided by these measures, "unpredictables" may be converted to "predictables".

As the case stands at the moment, an anomaly exists when interpreting the results of a nursing audit to establish restorative nursing quality. It is possible to achieve a good nursing audit whilst rendering a bad nursing service with poor patient outcomes. This is because neither the nursing audits nor the state and provincial healthcare inspections have any indices to evaluate patient-evidence outcomes in sub-acute or non-acute facilities. Ironically all such audits and inspections are primarily designed to verify that no harm is done to the patients and this is done with checklists to verify that the appropriate nursing documentation, processes and procedures are in place to prevent injury, infections, or any possible harm to the patient. Without any reference or motivation to increasing patient independence and only emphasising the prevention of any possible harm, the cautious nurse would rather keep the patient immobilised to achieve a good nursing audit.

1.5 Research purpose

The purpose of the study was to validate a set of new nursing measures to measure patient functionality. Such measures, if useful to nurses would provide accurate patient-evidence based data to the healthcare management and funding industries in South Africa to do extensive outcome analysis.

1.6 Research problem

The research problem is how to find a scientific method that would empirically validate both the usefulness and accuracy of the newly developed ALPHA, BETA, GAMMA and DELTA scales. As the researcher used experienced nurses as respondents to develop the four nursing scales, there is an expectation from the nursing participants that the scales already have significant nursing utility. Furthermore, the researcher also estimated high levels of accuracy as he related the item difficulties and patient abilities in a coherent and integrated way using

the nursing intuition from experienced nurses. However, both the nurses' expectations of nursing utility and the researcher's estimates of construct validity have not been scientifically validated. A scientific study needs to be done to prove these assumptions right or wrong. The entire study therefore follows a deductive approach to investigate these assumptions with the research question: How can the study best provide the scientific evidence required to conclude that the ALPHA, BETA, GAMMA and DELTA have both nursing utility and construct validity.

1.7 Research objectives

In line with the deductive approach of the study, two main objectives were identified. These were to test the assumptions of nursing utility and construct validity on each of the ALPHA, BETA, GAMMA and DELTA.

The first main objective, nursing utility was divided into three sub-objectives namely:

- to establish if the measures could be embedded into the nursing process e.g. scores done routinely on patients, recorded in the patient documentation and implemented in the nursing care plan;
- to create a uniform language amongst the nurses, multi-disciplinary team members and case-managers e.g. when discussing patient functionality status, goal setting and outcome; and
- to apply the data to establish quality of nursing e.g. routine calculations on patient outcome to infer nursing performance and quality assurance.

The second main objective, establishing construct validity were divided into two sub-objectives namely:

- to examine the validity (construct validity) of the nursing measures;
- to examine the reliability (internal consistency) of the nursing measures.

1.8 Scales, measures and numbers

Terminology when referring to “scales providing scores” and “measures providing measurements” needs clarification. When is a scale a measure and when is a score a measurement? Ben Wright and Mike Linacre announced the ground rules in 1989 with their classic article: *“Observations are always ordinal; measures however must be interval.”* Thus, a scale is an instrument, tool, test, or questionnaire based on observational interpretations of human functioning, behaviour or intellect. As such, a scale produces a score allocated by the ground rules to the specific scale. However, these “raw” scores remain observations, which are always of an ordinal nature, meaning they are unavoidably ambiguous. They are only qualitatively-ordered observations with a number attached. Statistically, score data have no precision and have no value for any further arithmetic inferences.

Measures, on the other hand, are of an interval nature, which provide well-defined linear measurement data with realistic precision and validity estimates useful for parametric analysis. Measures are fundamental in nature, (e.g. length, temperature, weight, etc.), whereas scales are not. Thus, on the issue of construct validity, this study will investigate methods to calibrate ordinal “scales providing scores” into interval “measures providing measurements”. At the onset of this study the ALPHA, BETA, GAMMA and DELTA tools are considered scales providing raw score data. It is anticipated that this study will reconstruct these scales into fundamental measures providing measurement data for statistical analyses. The terminology used in this study will thus evolve (e.g. scores becoming measurements) with the Rasch calibration process where the four ordinal scales will be converted into fundamental measures with linear interval characteristics.

1.9 Significance of the study

Currently, all measurements of patient functionality remains locked-up in the restorative nurse's intuitive knowledge. For as long as this knowledge remains immeasurable, the restorative nursing process remains unmanageable. If this study provides scientifically based nursing measures to unlock the restorative nursing sciences in South Africa, it may add significant value to the nursing profession, the patients requiring restorative care, and the emerging sub-acute and non-acute healthcare services in South Africa.

By adding a measurement dimension as a daily occurrence in the restorative nursing process, nurses could find themselves in a more assertive position where they could express themselves more accurately in relation to their assessments, diagnosis, techniques and evaluations. They could also find it easier to calculate nursing performance in terms of patient-based outcomes, to communicate quantifiable goal-directed nursing plans, and to recognise objective benchmarks that require change in approach, treatment or even discharge of patients. For the patient receiving restorative nursing, the measurements will indicate a new approach of facilitative nursing whereby the patient proactively participates in their own recovery process. For the sub- and non-acute health care management and funding industries the benefit of obtaining longitudinal routine actual patient-evidence data from the nursing profession is that this will unlock numerous barriers to enhance the funding of outcomes based services. Most importantly, such data will create a opportunity to find statistical methods of analysis whereby the unpredictable patients may in future be referred to as being predictable, given the appropriate nursing at the appropriate time for the appropriate reason.

The original contributions by the researcher to enable the restorative nursing profession to achieve the preferred patient-based outcomes as discussed above will be:

- a group of validated outcome measures to accurately assess and evaluate patient function;
- a new patient-level-method of analysing healthcare outcomes;

- a uniform language to discuss and describe patient functional gains firstly amongst nurses and secondly across a team of multidisciplinary members;
- a statistical research platform to analyse the value and performance of the nurse within the emerging sub-acute sciences;
- a triage framework for the nursing case-manager.

1.10 Definition of key concepts

For the purpose of this study and for the purpose of continuity, the following terms have been used:

- Fundamental measure

The distinctive attribute of a fundamental measure is the requirement for an arbitrary unit of difference that can be iterated successively along the latent variable; e.g. centimetres constituting length or grams constituting weight.

- Patient

For the purpose of this study a patient is a person who needs or receives restorative nursing care. Although numerous alternative terminologies are available (e.g. user, client, person, subject, beneficiary and more) due to concerns related to dignity, human rights and political and ethical correctness, for the sake of consistency in this thesis, the term patient will be used for any person requiring or receiving some form of restorative nursing care, be that in an acute hospital, sub-acute facility, home based care or long-term care environment.

- Rehabilitation

In this study the term “rehabilitation” refers to the comprehensive service rendered to both a sub- and non-acute patient with a residual motor, cognitive or behavioural functional loss due to illness, disease or trauma (e.g. stroke, multiple trauma, neurological diseases, mental illness, old age, disabilities etc). Although active rehabilitation is usually rendered in a sub-acute nursing setting,

but this service may also continue at home as a home based care service, or in a long-term care facility as a nursing program to maintain human function.

- Restorative nursing

It refers to the process whereby the nursing process' main aim is to increase or maintain human functionality in a sub- or non-acute nursing environment. A literature review on restorative nursing is available in Chapter Three.

- Scale

In general, the terms “scale, instrument, test, tool, questionnaire” may all refer to the concept of producing numerical scores to explain a phenomenon at an ordinal level (Bond & Fox 2007). However, each term brings its own approach and connotation to the general idea of collection of ordinal observational data as explained earlier in this chapter under the heading: Scales, Measures and Numbers. In this study, the term “scale” will be used collectively when referring to any of the above terms, but when a specific approach or connotation is better explained by the terms “scale, instrument, scale, test or tool”, these terms will be used interchangeably. However, none of these terms will refer to a “measure”, which identifies a higher level of measurement qualities, namely linear interval characteristics (Stevens 1946).

- Validation

The term “validation” in this study refers to two methodological studies that jointly and equally seek to validate the scales under discussion. The first investigation explored the nursing utility and the second investigation explored the construct validity of the scales. Nursing utility includes confirmation of whether the nursing measures could be generalised to other care settings in the real world of nursing and the measures' appropriateness and usefulness to be embedded routinely in the nursing care plans. Construct validity refers to the methodological investigation of how closely the nursing measures approximate a fundamental measure to produce measurements useable for parametric outcomes analysis.

- Unpredictable Outcome

The term “unpredictable outcome” refers to the outcome of a patient in a sub- and non-acute nursing care setting. The term “outcome” in this study refers to the quantifiable change in a patient’s functional levels from the admission to the discharge date. When the change indicates a functional gain, the outcome was considered a positive outcome; if a loss is evident, the outcome is reported as being negative. Sub- and non-acute care patient requiring restorative nursing has unpredictable outcomes as the nurses rendering their care had no means to measure their outcomes.

1.11 Ethical considerations

This study was approved by the Research Ethics Committee (Medical) of the University of the Witwatersrand and an ethical clearance certificate with the number M 10524 was obtained (Annexures B,C and D). The researcher accepted the responsibility for conducting the research in an ethical manner as detailed in the Democratic Nursing Association of South Africa’s document “Ethical Standards for Nursing Research” published in 2005. These standards were adhered to in its entirety. Although this study was entirely based on observational research, without any risk, harm or exploitation to the participants, consent was still obtained from the management of the nursing facilities involved. Annexures E and F are copies of such consent forms. As the research used scores from the nursing records, consent was not needed from the patients in this regard. Confidentiality and anonymity was ensured by preventing any linkages of the research data or thesis which could reveal the identity of either the participants (patients, nurses, therapists) or the facilities included in this study. The researcher undertook to conduct the research process with integrity; warrant the analysis to be trustworthy and valid, and certifies that the results as well as the recommendations were disseminated appropriately. Finally, the researcher declares that he was never employed by a nursing facility or in any other position of authority to influence the nursing participants during the qualitative studies.

Over and above the ethical considerations declared above, the researcher further accepts the responsibility for:

- informing participants regarding the purpose of the research before they were invited to participate;
- informing the participants that the group report will be shared with many, but the individual contributions will be anonymised;
- ensuring that participants of focus groups will not be subjected to any psychological harm due to over disclosure;
- maintaining an honest relationship with participants including not pressuring them to contribute;
- not allowing a participant to participate against their will;
- allowing a participant to make an individual decision or contribution without fear of favour or negative consequences.

1.12 Summary

In Chapter One the voice and vision of the definitive nurse, Florence Nightingale, was resonated. Her determination and steadfastness to use patient-based evidence as an outcomes measure for good nursing and healthcare reforms inspired this study. It is inspiring to see that her principles are so universal that they are still applicable to today's nursing processes. Therefore, with stating the current lack of validated empirical evidence in the restorative nursing process and the negative implications thereof for the nursing profession, the patient and healthcare reform, it was appropriate to revisit the Nightingale principles of the 19th century in an attempt to re-address the current situation. The research problem, aim and objectives of the study have been reported. In line with Nightingale's principles, the significance of the study's contribution to restorative nursing, to patients requiring restorative interventions, and to the sciences of restorative nursing has been tabled. In Chapter Two the considerations that guided the design and development of nursing measures will be discussed.

CHAPTER TWO: CONSIDERATIONS IN THE DESIGN AND DEVELOPMENT OF THE NURSING MEASURES

I make progress by having people around me who are smarter than I am... and listening to them.

Henry J Kaizer

2.1 Introduction

In Chapter One both the absence of and the need for nursing measures to track patient outcomes in the sub- and non-acute nursing settings have been discussed. In acute nursing settings adequate biometric measures underpins the nursing process in guiding patient outcomes. However in the sub- and non-acute nursing settings, where functional gains towards independence is the nursing objective, the restorative nursing process has no such instruments to guide patient outcomes. To overcome this situation the researcher facilitated the design and development of such tools using expert nursing respondents that have significant practical nursing experience in the sub-and non-acute nursing care settings.

In this chapter the researcher discloses the various considerations taken into account during the design and development of the four nursing measures, which conceptually contributed to the validation of the nursing measures. Based on expert nurses' intuitive knowledge, hypotheses and theories were used to formulate these considerations. In the later chapters the scientific validation studies will reveal whether the researcher's and nursing expert's concepts were sound or not. The design and development of each individual measure will not be discussed here but will be dealt with in Chapters Five to Eight with one chapter allocated to each of the four scales.

2.2 Considering who qualifies to design and develop the nursing measures

Due to the continuity of care of the nursing profession, the nurse as the primary caregiver is best placed amongst the healthcare professionals to develop intimate and emotionally intense relationships with patients. This inevitable nurse-patient relationship, linked with ongoing observations of patient functioning, forms the foundation of the nurse's intuitive knowledge which, according to [Billay, Myrick, Luhanga and Yonge \(2007\)](#), is a legitimate nursing resource. In fact, [Billay *et al* \(2007\)](#) emphasised that it is critical for researchers to recognise and explore the rich intuitive knowledge of nurses, and [Smith \(2007\)](#) requested from nurses to embrace the uniqueness of this clinical skill. [Gobet and Chassey \(2008\)](#) proposed a new theory of nursing expertise (problem solving) based on nursing intuition (perception).

For these considerations the researcher fulfilled the role of facilitator in the design and development of the nursing measures, and invited nurses to mine into their intuitive knowledge and provide the greater framework required, consideration in the development of these measures. They were selected for their rich practical experience in the clinical scenarios to be covered by this study. Although nurses knew intuitively what is workable and what information is useful to the nursing process, the researcher had to guide them through the other considerations that would result in fundamental measurements, as discussed below.

2.3 Considering the therapeutic contributions to the nursing measures

Patients in sub- and non-acute nursing environments are receiving restorative services which in most instances require a multidisciplinary approach. For this reason, it makes sense to take cognisance of the existing measures already in

use by the therapeutic professionals in this environment and to identify those language factors, constructs and barriers that exclude the nursing profession from using the same measures. If the measures could be adjusted to allow nursing participation, it would benefit team unity by facilitating a universal language across the healthcare professionals. Thus a strong consideration was not to re-invent a different structure, format or framework, but rather to retain that which is useful to the nursing profession and to adjust, edit, re-design or re-develop therapeutic concepts that do not fit the nursing sciences and that do not benefit the nursing process. Therefore, it was critical not to introduce measures that have the potential to divide and isolate the professionals, but to rather use those that serve as a medium to enhance the ties that originally brought the professions together – i.e. patient functionality.

2.4 Considering the context of restorative nursing practices

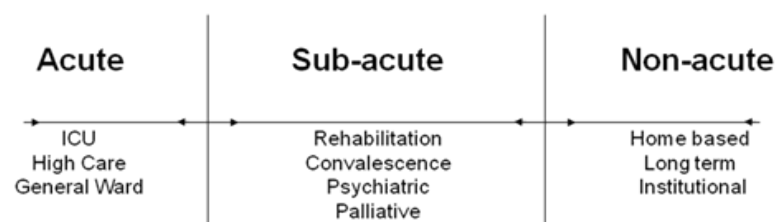
Over twenty years ago, [Eagar and Innes](#) (1992) published the watershed article that led to the sub- and non-acute nursing sciences in Australia. They hypothesised that the term “acute hospital” is a misnomer as few, if any, hospitals can be exclusively defined as acute. Patients are acute, hospitals are not. These authors stated that patients undergo distinct episodes of treatment in acute hospitals that can be classified as acute, sub-acute and non-acute, and that these episodes of treatment should take place in designated levels of care with distinctly different clinical objectives that must be taken into consideration when reporting on patient outcomes analysis. From a nursing perspective an episode of nursing care also refers to a goal directed set of nursing interventions and techniques to achieve an anticipated outcome.

The Eagar and Innes (1992) hypothesis was tested by two other studies in 1998. The first study by [Flintoff, Williams, Williams, Basinski, Blackstein and Naylor](#) (1998) was a large survey of 105 Canadian hospitals subjecting 13,242 in-patient files to an InterQual Criteria Utilisation Review. Their findings

concluded with the Eagar and Innes hypothesis. On admission, 62% of patients were acute, 19.7% were sub-acute and 18.7% were non-acute. On the subsequent day after admission 27.5% were acute, 40.2% were sub-acute and 32.3% were non-acute. The second study by [Weaver, Guihan, Hynes, Byck, Conrad and Demarkis \(1998\)](#) used 43 American acute hospitals and reviewed 858 patient files also using the InterQual Criteria Utilisation Review methodology. Overall their findings concurred with the Flintoft *et al* study but a significant secondary result was that 33% of post-surgical cases and 42% of post-medical cases require sub-acute care.

In 2004, [Loubser and Raath](#) did a cross-sectional study for the South African National Department of Health in 27 state hospitals using the South African Database for Functional Measures' instruments to score 5243 inpatients. The patients' needs for different treatment episodes were as follows: acute 34%, convalescence 43%, rehabilitation 9%, palliative 5%, and home-based care 10%. These findings not only concurred with the [Eagar and Innes \(1992\)](#) hypothesis, but it also focussed the Department of Health's attention on the significant need for restorative nursing interventions to be included in the healthcare continuum.

According to [Eagar and Innes \(1992\)](#), these distinctive interventions of nursing practice also require distinct settings of patient care. Taking into consideration the realities of the continuum of nursing care, the researcher decided to identify each restorative nursing episode and development measures to reflect the objectives of the restorative nursing episode. This is illustrated in [Figure 2.1](#).



[Figure 2.1](#): The continuum of acute and restorative nursing episodes and settings.

2.5 Considering the interventions expected from restorative nursing

According to the Australian National Sub- and Non-Acute Patient (AN-SNAP) Case-mix Report (Eagar, Gordon, Hodkinson, Green, Eagar, Elven *et al* 2001), the aim of restorative nursing in sub-acute care is to maximise recovering of human function in the shortest period of time. The researcher added the objective of quantifying human function, and thereby explaining the outcomes of the restorative nursing process for the patients. The quantification of function included a baseline of patient functional status on admission, monitoring of progress throughout treatment and predicting the point of maximal restorative improvement the patient can achieve in the sub-acute care setting.

In sub-acute care, it is generally accepted that the Activities of Daily Living (ADLs) such as self-care, sphincter control, transfers, locomotion and cognitive abilities fall within the scope of the restorative process. Activities of daily living are specifically sensitive and helpful in explaining outcomes of rehabilitation, convalescent and palliative care patients, specifically in their end-of-life phase (Daniels 2004). However, in mental healthcare patients the ADLs are not sensitive enough to accurately reflect the functional changes as observed in this group of patients. These patients primarily experience a loss in their executive functions, which present with a variety of specific sets of clinical and behavioural syndromes (e.g. hallucinations, delusions, obsessional behaviour, loss of focus, etc) observable by psychiatric nurses. With clinical improvement these patients regain control of their executive functions and as their clinical and behavioural symptoms disperse, the nurses can observe how their functionality changes with improvement or decline.

In non-acute care settings, such as home-based care, assisted living and long-term care where services are rendered to geriatric communities, people with

intellectual disabilities, people living with disabilities or chronic illness, or those who require short term support at home, the patient's ability (functionality) to live independently is a vital requirement. The patient's ability to continue living independently not only becomes the nursing objective in this care setting, but the techniques to achieve independent living also become the nursing science of preference (Graf 2007). In this context the Lawton's Instrumental Activities of Daily Living (IADLs) scale has become synonymous with independent living abilities (Lawton & Brodie 1969; Graf 2007).

Thus, when designing measures for the sub-acute process, the researcher had to consider the various ranges of functional gains or losses anticipated within each distinct episode of nursing care. Although the change in function remained the overall construct of measure, the range of functional change as well as the resultant nursing sciences to manage the specific change, would differ between the nursing episodes on the continuum of care. Figure 2.2 illustrates the scope of functional changes in sub-acute restorative care.

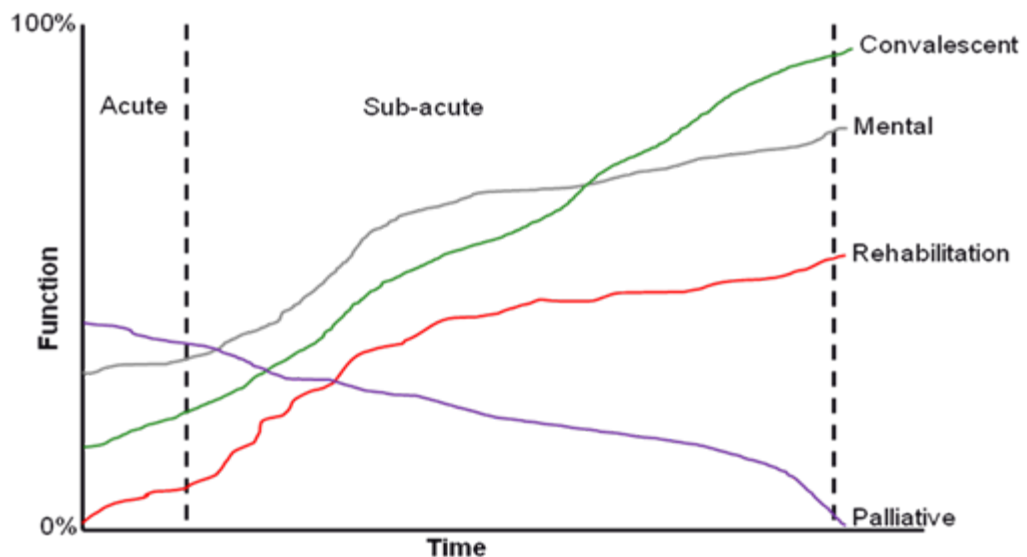


Figure 2.2: Scope of functional changes in sub-acute restorative care.

2.6 Considering the interface between acute nursing and restorative nursing

There is general consensus in the literature that the acuity of a disease, illness or trauma is defined by four criteria: rapid onset, a short duration, impairment of normal functioning and urgency for prompt support and treatment ([Mosby's Medical Dictionary 2009](#)). Acute, curative nursing requires a mix of sophisticated clinical skills and technology to preserve life by supporting failing organs and systems, performing investigative procedures, implementing and managing treatment programs and supporting vital functions. Typically, during the acute nursing process the patient is immobilised and the nurse is continuously on alert to observe, record and communicate all warning signals that might indicate a complicated recovery ([Daniels 2004](#)). The main objective is to get all the defunct bodily organs and systems back to functioning normally. When this turn-around has been achieved the patient is considered stable, meaning the patient's vital organs and systems are functioning independently and the patient is out of the life threatening zone. This state signals the end of the acute nursing process and the beginning of the restorative nursing process.

In the restorative nursing process the focus is on regaining the residual activities of daily living functional loss caused by the acute episode. The [AN-SNAP Report of 2001](#) classifies a sub-acute episode of nursing as one that is:

- 'provided for a person with an impairment, disability or handicap' and
- 'for whom the primary treatment goal is improvement in functional status' and
- 'which is evidenced by an individualised and documented initial and periodic assessment of functional ability by the use of a recognised functional assessment measure' and finally
- 'an individualised multidisciplinary rehabilitation plan which includes negotiated functional goals and indicative time frames'.

Australia's [Victorian Report \(2001\)](#) explored the interface between acute and sub-acute nursing episodes of care. It highlighted the importance of transferring patients at the optimal time from acute care to sub-acute care for the start of the restorative nursing process. Transferring the patient at the right time has significant benefits for the patient and for the nursing process. Outcomes for patients are better when formal rehabilitation commences earlier and there may also be improvements in overall length of hospital stay and cost of care. Conversely, there may be adverse outcomes when patients are transferred too early. For example, patients who remain medically unstable may not be able to be safely managed in the sub-acute facility, as unstable medical conditions could render the rehabilitation process less effective, and undue time could be wasted if the patient has to be transferred back to the acute care facility, or other centre, for diagnostic or medical evaluation. Typically, older patients may, after an acute episode, find themselves with multiple co-morbidities or general debility, and while they no longer seem to require acute nursing as their organs and systems are functioning within the normal variances, they often require a period of restorative nursing.

With the above considerations in mind the researcher had to contemplate nursing measures that would assist in managing the acute / sub-acute interface. As the acute nursing episode is defined by the extent of non-functioning of the human organs and systems, a measure for the interface must establish the extent to which the organs and system functioning has returned to normality - or "stability" as it is known in the clinical environment. The need is to objectively quantify if the patient's organs and systems are responsive enough to start restorative nursing.

There are two approaches to consider for measuring the functioning of organs and systems. Firstly, this can happen through the clinical evaluation of quantitative diagnostic analysis provided by the vital signs, laboratory reports, radiological reports etc. Few, if any, of these are nursing instruments. Although

the final assessment to transfer the patient for restorative nursing lies with the physician's interpretation and judgement of the acuity level, this is often done in consultation with the nurse.

The second approach to consider is the relationship between the acute nursing burden of care (and interventions) and the functioning of the organs and systems. The assumption is that the lower the independent functioning of the organs and systems, the higher the acute nursing burden of care. This approach to classify acuity was already mentioned in Chapter One and is used extensively in South African private hospitals to group patients broadly in the same resource utilisation cost groupings (RUG) and is taken into consideration when transferring patients from ICU to high care and to the general ward. The interfaces between these three acuity levels are clearly demarcated by the acute nursing techniques applied to support failing organs and systems. What is, not clear however, is the interface between general ward and sub-acute care and the clarity on when a patient's organs and systems are functioning independently enough to be transferred to a nursing environment where the focus is on activities of daily living. Thus, the consideration for the researcher was to design and develop a nursing measure whereby the nursing profession could quantify the level of independent functioning of the bodily organs and systems to establish "stability".

2.7 Considering nursing utility

Certainly the most significant consideration to keep in mind is the impact the measures might have on nursing utility, which can be best described as the measures' usefulness in the nursing process. Without the nursing profession acceptance of the measures as being useful to their day-to-day practice, no routine data will be collected, and if nurses are mandated to collect data from any authority the data could be inaccurate. Routinely collecting data that are not

considered to improve nursing quality will not be treated with the care and respect required. The extent of utility phenomena is multi-dimensional. A literature review highlights the following inquiries and highlights some of the variables involved:

- *Is the measurement relevant to the patient and the assessment situation (Innes & Striker 2003b)?*
- *Is the measure clinically meaningful in meeting the needs of the patient, the nursing process and the funder (Innes & Striker 2003)?*
- *Is the measure comprehensive in meeting the scope of patient function (James & Mackenzie 2009)?*
- *Is the measure accurate and does it provides the correct information needed to monitor the patient's functional ability (James & Mackenzie 2009)?*
- *Is the measure flexible enough to be applied in various nursing episodes (Toomey, Nicholson & Carswell 1995)?*
- *Is the measure practical enough to be applied, administered and interpreted with ease (Gibson & Strong 1997; Simmonds 2002)?*
- *Is the measure cost effective (Gibson & Strong 1997; Simmonds 2002)?*
- *Does the measure's credibility emphasise the observers' experience, skills, knowledge, and training (Innes & Striker 2003)?*
- *Is the measure suitable for its intended purpose (Gibson & Strong 1997)?*
- *Can the measure realistically be completed in full every time (Innes & Striker 2003b)?*
- *Does the measure provide organisational information to be implemented (Innes & Striker 2003b)?*
- *Is the measure providing valuable insights to the patient and the attending nurse (Barbara & Whiteford 2005)?*
- *Is the measure adaptable enough to be used on various disabilities and situations (Gibson & Strong 2002)?*

In addition to the broad clinical utility questions posed above, the following specific administrative and managerial benefits to the nursing process must also be considered:

- *Can the measurements be done through observations only?*
- *Can the measurements be done routinely (daily)?*
- *Can the measurements be imbedded into the patient records?*
- *Can the measures be used to monitor patient progress?*
- *Can the measurements be used to quantify nursing performance?*

Still within the nursing utility realm, a concerted consideration was also given to who the eventual rater of the measure would be, as the measures had to fit into their day-to-day operational framework. When it was established that the nursing assistants are the most appropriately placed to do routine direct observations of the ADLs or the IADLs, those measures were designed to best fit the difficulty level of the nursing assistants and their specific frame of reference, allowing them to easily familiarise the measure within their current practice. Simultaneously, added values were considered, such as allowing the primary nursing caregiver to adhere to the various theoretical levels of the nursing process.

2.8 Considering construct validity

As the intent of the researcher was to subject the four scales to the Rasch measurement model (RMM) for validation purposes, the design and development of each scale had to conform to the four basic requirements any scale has to comply with before the RMM can be applied. This includes the raw score data to conform to uni-dimensionality, local independence, monotonicity of the latent trait and invariance of the data ([Osborn 2008](#)).

Thus, to increase the probabilities of fitting the RMM and to simplify the development, the researcher based all the nursing measures on the same construct methodology. Firstly, to comply with the one-dimensional consideration, a single domain (trait) for each scale was identified, then the items (latent traits) that will support the selected domain were considered. Each item must make an equal but different contribution to the underlying latent trait. Care must be taken that that no two or more items contribute similar information. Finally, each of the items was divided into seven hierarchically ordered categories, whereby the first category represents the lowest possible score (lowest patient ability on the item) and the seventh category represents the highest possible score (highest patient ability on the item). The categories two to six represent the hierarchical order (monotonicity of the latent variable) of observations of patient functional gain or decline. Each scale item therefore represents a singular difficulty attribute (latent trait) to the rating scale domain (underlying trait) and each item is divided into seven observable levels in terms of patient functional abilities to perform the item difficulty (See Figure 2.3).

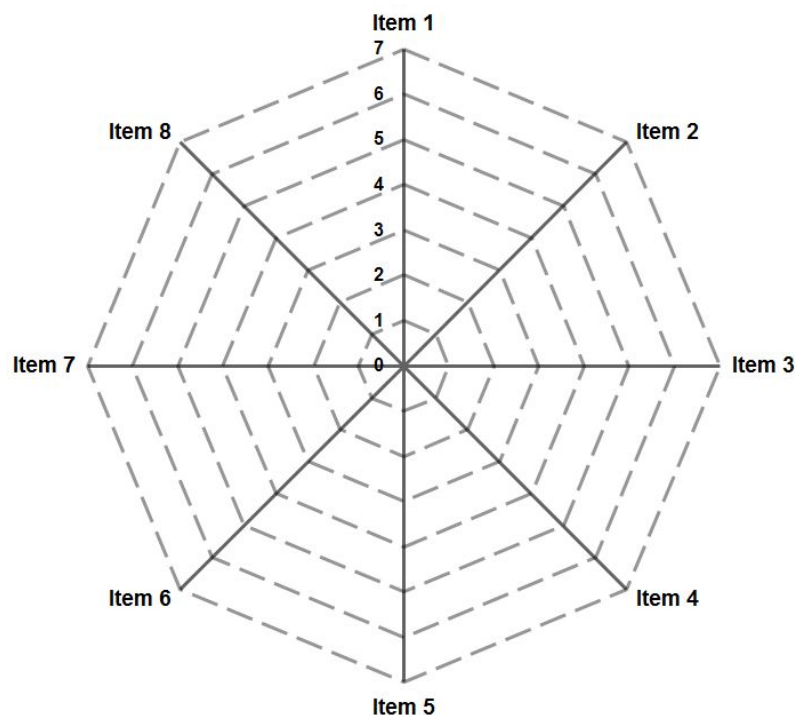


Figure 2.3: Radar graph example of measure constructs.

For RMM validation purposes, the assumption is that each of the four nursing measures has its own unique clinically sound underlying trait (domain) for scoring, each having its own unique set of latent traits (items) each contributing equally to the construct of the scale, and the latent traits (Items) having its own set of latent variables (observable categories) well modelled hierarchically to provide as close as possible linear interval level data on the latent trait. When involved in the measurement of human functionality the likelihood that this hypothesis will hold with a validation rating of 100% is extraordinary. To establish what constitute a perfect human measure, [Georg Rasch \(1960\)](#) has developed a mathematically perfect measurement model, and when ordinal scores of newly developed scale are subjected to the RMM analyses the outcome will reflect how closely the new scale “fit” the perfect RMM.

2.9 Considering routine measures

As the scales had to deliver longitudinal scores on tracking patient functional changes as they occur, they had to function as objective observational scales. Thus to be successful, these scales had to be accepted and included into the mainstay of vital nursing measures (e.g. thermometer, baumanometer, input/output measuring and charting) and observations (e.g. clinical signs and symptoms). If this can be achieved, nurses by virtue of their practice, will score and report on patient functional outcomes daily. If a scale can attain the level of vital statistic consideration in the restorative nursing process, then there is a real expectation that routine longitudinal functional data would be forthcoming from the nursing profession. The collateral benefits of routinely produced information are the ever evolving levels of accuracy achieved as the nursing skills develop as well as producing longitudinal and actual data on patient functioning as it changes and as it is observed. There is anticipation that the nurses would finally include functional measurements into their assessments and evaluations.

2.10 Considering burden to apply, time and efficiency

Although mentioned low on this list of considerations, burden and time to apply, is most probably first on the list of nursing considerations. Measures may be useful to the nursing process as mentioned above in nursing utility, but when a useful nursing measure require additional nursing time in an overloaded nursing process, such measures are at risk of being summarily rejected by the nurses. On the other hand, a measure producing relief to the overburdened nurse and generating additional time and space to improve quality of nursing might well be welcomed and favourably considered.

2.11 Considering sustainability and focus on the end-user

Developing nursing measures for data collection and stakeholder analysis require sustainable processes to ensure the collection of quality data. The basic considerations to generate sustainability are integral to the design and development, and the following processes were considered:

Firstly, identify who is the preferred end-user nurse of the specific measure. Thereafter, ensure that the measure does not exceed the difficulty level of the end-user and is written in the day-to-day language used amongst their peers. With this in mind, develop a training manual that is within the particular nurse's scope of practice and could be used as a guide and reference for a particular level of nurse. Also, develop testing material (e.g. case studies) to rate competency and set accrediting criteria for accepting the end-user data into the system, including regular and ongoing re-testing programs and facility credentialing programs. By taking the measure to the end-user, allowing them to identify their daily nursing tasks within the measurable observations should theoretically create ownership by the nurse and thereby reduce the risk of the nurse rejecting implementation of the scale.

2.12 Considering secondary use for healthcare reform

The primary consideration for the clinical need of the nursing measures was discussed extensively in Chapter One. However, a secondary case can be made out for the South African healthcare funders and providers who are high level stake holders in the healthcare system. They have an interest in outcomes measures and their needs must be taken into consideration in the design and development of the measures.

Over the last decade the South African healthcare landscape has changed into a market-orientated commerce where funding industries such as the manage-care organisations, medical schemes, and insurance and re-insurance companies have become bulk corporate buyers of healthcare services on behalf of their clients or membership. In the drive to increase business through membership, outcomes focus has shifted from “how care was provided”, to make provision for “how care was experienced” by clients and members. This phenomenon requires funders to enquire about valid patient–evidence based data. Moreover, this new dimension places the spotlight on healthcare providers to quantify the patient outcomes achieved by their services, rather than quantifying the resources they have put into their services ([British Department of Health 1998](#)).

Furthermore, as the new bulk purchasers of healthcare services become cost driven, they require assurances that the services purchased are both clinically effective and cost-effective – thus driving the “outcomes agenda”. This places a burden on sub- and non-acute providers to render proof of the efficacy and efficiency of their practices. Healthcare clinicians should not only provide measureable clinical goals, but also provide assurances with existing performance data that they are capable of achieving those goals consistently and appropriately. Funders are looking for best practices for their members/clients. Although many argue that big business and/or fiscal restraints

might drive the outcomes agenda, [Unsworth \(2000\)](#) states that it cannot be argued that this does not go hand-in-hand with the humanitarian desire to provide the best quality healthcare services to the patient.

As the outcomes agenda increased the pressure on healthcare professionals to render proof of their performance, so did the providers start the “total quality management” debates of health service delivery. This included “critical appraisals”, “reflective practices”, “systematic audits”, “best value reviews”, “service evaluations” and more ([Richards 2002](#)).

However, for this study the most significant consideration from this debate is the “evidence-based practice” methodology, which [Straus, Richardson, Glasziou, Haynes et al \(2005\)](#) referred to as the process of ongoing gathering of knowledge from practice, tested in research, and which continues to inform all stakeholders how to best contribute their efforts towards achieving excellence in healthcare services. [Peile \(2004\)](#) described these core-activities at the root of evidence-based practice as a questioning approach to practice, leading to scientific experimentation, meticulous observation, enumeration, and analysis replacing anecdotal case description, recording and cataloguing the evidence for systematic retrieval. The classic work of [Cochrane \(1972\)](#) suggested that because resources would always be limited, it should be used to provide forms of health care which had been shown in properly designed evaluations to be effective. Cochrane maintained that the most reliable evidence is that which comes from randomised controlled trials (RCTs) ([Straus et al 2005](#)). However, [Dale \(2005\)](#) argued that different forms of research, other than randomised controlled trials, are valid and in many cases more applicable to nursing practice, and that nurses need to determine what constitutes relevant and best evidence for the profession. As the matter stands currently, the restorative nurses are not capable of participating in evidence-based practice research as anecdotal case description evidence is still collected as evidence of patient independence.

According to the South African Council of Medical Schemes ([Van der Merwe 2009](#)) evidence-based management is one of the fundamental activities of clinical governance. The routine observation, collection, recording and charting of valid patient evidence-based functional outcomes data is the natural starting point of a clinical governance framework. If all such data is entered into a multi-facility clinical governance database, statistical and actuarial analysis will unlock patient-based benchmarks and evaluate trends that are currently unavailable as they are considered indefinable and nebulous. Reporting on such clinical governance activities will provide the service provider and funder organisations with new insights in accountability, which will assist in continuously improving the quality of services and creating an environment whereby excellence in clinical outcome will flourish. Clinical governance reporting can thus be extended to reveal best nursing clinical guidelines that define best nursing practices, resulting in quality nursing services to the patient.

Therefore, when sub- and non-acute clinical governance inferences are dependent on the availability and validity of nursing measures, the considerations for the design and development of such measures must conform to the expectations of multiple and opposing stake-holders to deliver objective and accurate data. If such data are imported into large-scale, multi-facility databases for clinical governance outcomes reporting, high-stakes analyses are involved. The measurements should be of high quality in order for the inferences based on these results to be valid, and for the decisions based on these reports to be useful to all stakeholders. The healthcare system has numerous levels of decision makers and the reporting on patient outcomes will be used by those that have a legitimate right to access patient, nursing, facility and national outcomes reports.

Thus, when basic patient-based observational data becomes available through the nursing process, modern information technology is advanced enough to support large-scale data platforms of enormous power to analyse and resolve complex clinical governance accountability and risk taking issues, provided that

the basic nursing data is available and accurate. Whilst considering all these secondary interest groups, the researcher's considerations were to keep the design and development a basic and accurate representation of reality, allowing for future high level statistical and actuarial inferences.

2.13 Summary

In Chapter Two an overview of the considerations for the design and development to enhance the later validation of the nursing measures was discussed. Although the need for nurses to participate in the development of nursing measures was strongly emphasised, existing measures used by the therapeutic professionals should also be considered for possible contributions to nursing measures. The continuum of sub- and non-acute restorative nursing was highlighted for its diversity. Important considerations were that in its diversity, each nursing episode is seeking a different functional range representative of its patient outcomes. Furthermore, both nursing utility and construct validity were discussed as constant considerations when designing useful and accurate measures. Finally, secondary considerations for the use of longitudinal accurate patient based data were contextualised. In Chapter Three a literature review will focus on three concepts, namely the standing of restorative nursing sciences and processes, the current understanding and extent of nursing utility, and exploration of the Rasch measurement model as a new method to analyse construct validity in the health sciences.

CHAPTER THREE: LITERATURE REVIEW

"You shall not have in your bag two kinds of weights, a large and a small. You shall not have in your house two kinds of measures, a large and a small. A full and just weight you shall have a full and just measure you shall have."

Deuteronomy: 25 13-15

3.1 Introduction

This chapter incorporates a review of the literature on the three core subject matter that supports this study; namely the restorative nursing model, the concept of nursing utility and the RMM as a relatively new concept in the health sciences to determine construct validity. Firstly, the model of restorative nursing was explored for its origins, its conceptual framework, its ability to be measurable, implementable and seeking for evidence in the literature for any existing patient-centred nursing measures used routinely as a method to explore patient and nursing performance. Secondly, this review explored matters concerning the utility of routine nursing measures for restorative nursing, and finally, the mathematical techniques Georg Rasch popularised as a model for validating measures for human functioning with ordinal scales. A literature search for various methodologies and techniques to validate the clinical utility and the measurement properties of the proposed Alpha, Beta, Gamma and Delta nursing measures was conducted. For these reviews the PubMed, Ebsco: CINAHL (plus with Full Text) and the JSTOR (archival) subject databases were used.

3.2 Restorative nursing model

In the 1970's a popular notion was that long term care nursing facilities were

“warehouses” for frail elderly and disabled persons until they died (REF). The science of long-term care nursing was about creating comfort and dignity during this waiting period for the final day to arrive. The nursing process did not expect any participation from the residents and their inability to function was forced into total dependence by the smothering nursing care provided. In the 1980’s there was awareness that the traditional nursing practice not only deprived the elderly and disabled from their quality of life, dignity and independence, but also incurred high costs and was indeed a breeding ground for low staff morale. Nursing became system compliant, not patient compliant.

At this stage the nursing historians requested a relook at the original work of two nursing pioneers. Firstly the works of Sr Elizabeth Kenny were reviewed. She was an unorthodox, outspoken and controversial Australian, who in the early 1940’s, refused to treat her polio patients according to the standard practices of the time. She developed her own terminology, theories and practices for the treatment of the affected limbs of acute polio patients. Her nursing techniques of early active mobilisation directly opposed the global standard procedures of immediate immobilisation in splints for long periods. When she presented her nursing theories in the USA, her nursing practice was vehemently opposed by both the medical and political fraternities in the United States, but Sr Kenny was a maverick and she persisted with her restorative nursing care practise. Unfortunately she only resorted to her theories and techniques to convince her audiences, unlike Florence Nightingale who believed that patient-evidence based outcomes data was the essence of reform in healthcare. However, in 1942 the Times Magazine came to Sr Kenny’s rescue by noting an 80% recovery in polio patients undergoing her nursing techniques. Finally [Robert Bingham MD \(1943\)](#) reported on the patient outcomes of Sr Kenny’s restorative nursing programs. He concluded:

“Patients receiving the Kenny treatment are more comfortable, have better general health and nutrition, are more receptive to muscle training, have a superior morale, require a shorter period of bed rest and hospital care, and seem to have less residual paralysis and deformity than patients treated by

older conventional methods. The Kenny treatment is the method of choice for the acute stage of infantile paralysis” (Bingham 1943).

Sr Kenny passed away in 1952, but the Sister Kenny Institute based in Indianapolis still continues today with her legacy of patient advocacy through restorative nursing care (Acello 2009). The physiotherapy sciences also claim her legacy as an important originator of their practice.

The second pioneer, Sr Verah McIver rebelled in 1967 against the custodial care system of elder and disabled persons describing it as nursing methods that triggers a cascade of effects leading to a loss of pride and dignity and the death of the human spirit. She created an ability-enhancing nursing model through training programs for nursing assistance and management and supervision programs for the registered nurses in charge. She called it the Restorative Nursing Care Model, defined it as an enabling nursing process aimed at promoting physical and personal independence to restore the dignity and wholeness of the elder or disabled person (Mantle & Funke-Furber 2003).

Motivated by the Kenny and McIver legacies the Omnibus Budget Reconciliation Act of 1987 (OBRA 1987) initiated a system of re-imbusement for implementing restorative nursing principles into the long-term nursing care facilities in the United States of America. The OBRA Act revolutionised the process of nursing care for elder and disabled persons as the Act defined the care to “attain and maintain the residents’ highest practicable physical, mental, and psychosocial wellbeing.” Thus restorative nursing, the missing link between therapy and nursing, became mandatory in USA nursing facilities in 1987 (Wiener, Freiman & Brown 2007). The OBRA Act, also referred to as the “Nursing Facilities Reformed Bill” (1987), further mandated a nursing aid registry and training program based on the restorative principles.

The researcher found little, if any, evidence of any formal restorative nursing practices in the South Africa nursing sciences. It seems as if the traditional curative caring model and its audit processes are simply replicated into the sub- and non-acute settings, without taking cognisance of patient needs.

The opportunities restorative nursing avail to both the patient and the profession's outcome have not yet been considered. Although the curative nursing model remains the mainstay, without considering the restorative nursing sciences, unintentional harm is done to patients that require restorative care. To date, there is no evidence of any attempt that the nursing process in South Africa is investigating the opportunity to include physical and mental restoration into the nursing care process.

In the literature the restorative nursing care philosophy has only been reported within long-term nursing facilities caring for the frail elder and disabled patients. However, the researcher would like to facilitate the endorsement of this philosophy in various other spheres of nursing such as rehabilitation, mental health and even acute care when treating trauma and complex medical cases. All these nursing services are rendered to the unpredictable grouping of patients who require a comprehensive nursing program to regain their physical, mental and psychological functions. [Acello \(2009\)](#) concurs with the researcher in that restorative nursing care should be the preferred nursing process for all patients subjected to long periods of nursing services, irrespective whether in an acute, sub-acute, home-based or institutional care setting. The reason why the greater majority of patients require extended periods of nursing services is a result of functional loss, and according to [Acello \(2009\)](#) the benefits of applying restorative nursing philosophies to these patients are:

- It focuses the nursing process on the patient's functional needs which, in this context, is more of a consideration than patient diagnosis (e.g. with restorative nursing the core focus should be on overcoming and living with the hemiplegia, not treating the stroke).
- It focuses and addresses all the patients' functional needs across the nursing continuum as it plays itself out during the nursing care.
- It is goal orientated, explorative and understanding of how one weak area of functioning can affect the whole person.

- It increases independence and decreases dependence by helping the patient to attain optimum levels of physical, mental and psychosocial functioning.
- It helps the patient to regain lost skills, or master a new way of utilising those skills lost due to illness or injury, or adapt to life with limitations imposed by the impairment through assistive devices.
- It has to be verified by documentation which must be measurable.
- It teaches patient outcomes to family and staff members.
- It prevents complications resulting from inability, promote safety, and at a minimum, maintains current ability.
- It improves the patient's quality of life and self esteem.
- It is an extension of the basic nursing process rendered by registered nurses, nursing assistants and even unregistered primary caregivers.
- It is planned, implemented, and supervised by nursing professionals with assistance from other departments, if relevant to the nature of the nursing program rendered.
- It is integrated into regular nursing care and used and adapted where and whenever it is needed.
- Orders for the restorative program are written by the supervising nurse in-charge, no physician order is required.
- Therapists are consultants who should not write nursing orders (Acello 2009).

Very similar to the South African scenario, [Resnick \(2004\)](#) reported that Nursing Assistants (NAs) in the USA nursing led facilities provide up to 90% of the direct care and functional assistance to patients. Since 1987 it became mandatory in the USA to train long-term care nurses in restorative nursing practices ([OBRA 1987](#)). Yet, 22 years later [Resnick, Cayo and Pretzer-Abcoff \(2009\)](#) reported NAs' training in restorative nursing techniques to be still limited when compared to the formal training rendered to Professional Registered Nurses (PRNs). Clearly, the basic principles of restorative nursing have not reached the

coalface where it matters most. Similar to South Africa, [Resnick \(2004\)](#) stated that many nursing assistants use English as a second language, are working in a cultural environment that differs from their own and have benefitted little from the traditional lecture formats used in in-service training programs. As a result, many over time have learned ways of working that meet the facility demands for efficiency (e.g. getting the job done) at the cost of the patient's functional independence. [Resnick's \(2004\)](#) study in the USA concurs with the researcher's observations in the South African sub-acute and long-term care facilities.

In the literature numerous references are made to inadequate curriculums and training of NA's. As patients' needs vary extensively, so should NAs be trained to identify existing abilities and implement restorative nursing methods in their nursing process to enhance or strengthen the patient's independence? Indeed, this should be the prime objective. In order to achieve this objective, in-service training and education is required ([Nakhnikian, Wilner & Hurd 2002](#)) on how to support the complex patient needs such as the feeding of cognitively impaired patients ([Chang & Lin 2005](#)), dressing ([Engelman, Mathews & Altus 2002](#)) or improving communication skills ([Winchester 2003](#)) and how to deal with behavioural disorders ([Blair & Glaister 2005](#)).

The OBRA 1987 Act was a direct result of concern expressed in the USA about the quality of nursing practice in nursing facilities for the elderly and disabled persons. The Act required the development of a standardised assessment. As a result the Minimal Data Set (MDS) was developed and implemented in 1998 to offer a comprehensive assessment of nursing home residents, also known as the Resident Assessment Instrument (RAI-MDS). All nursing facilities in the USA caring for the elder and disabled persons were mandated to implement the RAI-MDS and provide data to a central database for the analysis of the quality of nursing care based on patient-based data. Reimbursement from the USA social security system was dependent on the provision of the patient-based data ([Mukamel & Spector 2003](#)). The MDS is a 284-item instrument divided into 15 sections to evaluate the medical, mental and social characteristics of nursing home residents. Assessments are made on admission and with quarterly intervals thereafter ([CMS 2010](#)).

The RIA-MDS provides significant advantages and disadvantages to reflect on when designing and validating nursing measures for patients with unpredictable outcomes. The disadvantages will be discussed under clinical utility (section 3.3) below. However, a major advantage is the valuable source of rich research data and information produced in terms of the quality of care provided in long term care settings for the elderly and disabled patients (Shin & Scherer 2009). Another advantage is that nursing staff is required to assess and reassess their patients on a regular and continuous basis (Hendrix, Sakauye, Karabatsos, Daigle *et al* 2003). As assessments on admission and follow-up quarterly assessments are mandatory, nurses are forced to plan, set-up and monitor the restorative process (Rantz, Petroski, Madsen, Mehr, Popejoy & Hicks 2000). As a result, there seemed to be a reduction in the use of physical restraints, dehydration and a marked increase of physical and cognitive function. The secondary analysis of the data also seemed to create quality indicators adding value to the nursing processes as it now becomes possible to do peer review reporting amongst facilities with similar patient profiles (Zimmerman 2003). The MDS data pool contains individual patient data which has significant value as longitudinal profiles of patient's functional, clinical and psychosocial decline become available and risk-adjusted health outcomes can be evaluated across nursing facilities (Mukamel & Spector 2003). With the above data and analysis available on all the USA nursing facilities, it is an administrative task to credential successful nursing facilities (Mor, Berg, Angelelli, Giford, Morris, Moore 2003).

In this study the hypothesis is that any measure collecting numeric data, routinely or periodically based on structured observations, can achieve the positive results and benefits discussed above. However, in developing such a standardised measurement instrument the real challenge is in achieving acceptance from the nursing profession to embed the data collection process as an integral part of the nursing process. The measures will not succeed if the nurses consider it meaningless to the nursing process. It will be rejected as an added burden to the RNs or the NAs.

3.3 Nursing utility

As early as 1995, Toomey, Nicholson and Carswell defined the utility of an instrument as the degree of conviction the users have about the usefulness of the instrument in their practice. Moreover, in 2005 Barbara and Whiteford declared it was also a useful method to validate the instrument as being appropriate for the purpose it was designed for as it addresses application practicalities such as relevance, suitability, feasibility, accuracy, comprehensiveness, credibility, flexibility, value and adaptability. Therefore, both the psychometric properties and usage issues should be addressed when validating a measurement tool as these two characteristics are complementary and interdependent on one another (James & Mackenzie 2009). The nursing utility of an instrument could be considered at three different levels of influence; at instrument level, at individual and organisational level (Wind, Goutteborge, Kuijer, Sluiter & Frings-Dresen 2006). At instrument level the ease whereby a measure can be embedded into the nursing process as a routine measure must be analysed; at individual level the usefulness, purpose and relevance of patient data collected must be assessed and at organisational level the clinical utility relates to the appropriateness and usefulness of the outcomes reporting provided by the data.

Numerous studies report on methods to test a instruments attributes of excellence. Innes and Straker (2003) used qualitative investigations to explore how therapists perceived the use of measurements in their day-to-day practice; similarly Barbara and Whiteford (2005) established qualitative methods through interviews of testing the perception of users of specific instruments. Others again used cross sectional quantitative studies with structured questionnaires distributed to health workers using the measurement in their practices (James, Mackenzie & Higginbotham 2007; James & Mackenzie 2009).

The nursing fraternity will only welcome a nursing measure if it makes clinical sense and serves as a catalyst to ease the nursing burden of care whilst simultaneously increasing patient and nursing outcomes. However, when a

nursing measure is perceived to be a managerial directive misusing nursing staff as data collectors of patient-based operational data, it is viewed as an added burden to the nursing process with a threat to the data integrity. The Resident Assessment Instrument – Minimal Data Set (RIA-MDS) with its 284-items divided into 15 sections is an example of the latter and remains in use only because nursing reimbursement is dependent on rendering the data.

A nursing measure based on observing patient progress using 284 items and 15 sections makes it too laborious and evidently impossible to implement routinely. For this reason the RIA-MDS is used periodically, collecting data only on admission and quarterly thereafter. The periodic collection of data makes no clinical sense to either patient or nursing outcomes as a patient's health and functional status change on a daily basis. Thus the reliability of RIA-MDS data is under suspicion (Blaum, O'Neill, Clements *et al* 1997), as it fails to collect longitudinal data explaining the variances in outcomes between the different patient case-mixes over time. These phenomena can only be explained through the analysis of longitudinal data collected routinely as changes occur. Periodic assessments do not add value to neither patient nor nursing outcomes. Furthermore, RIA-MDS measure is too extensive; the burden of application is too high, and the recording intervals too far apart to be useful as a nursing measure. With these poor nursing utility characteristics the RIA-MDS data is predicted to provide inaccurate data (Shin & Scherer 2009).

To overcome the negative nursing utility experienced by the RIA-MDS, the USA Social Security Agency appointed external clinical auditors to do the quarterly RIA-MDS assessments in the hope to collect accurate data. The problem however escalated significantly. Ethically it cannot be expected of the nursing profession to accept the external auditors to collect nursing operational data from their patients in facilities during their nursing watch. Furthermore, the external auditors are in most cases not the primary patient carers and not in a position to do direct observations. The auditors may be from various clinical disciplines, and as they also do not have the time to observe, comprehend and contextualise the patient's true situation quarterly, they rely on the resident NAs as proxy raters. The auditor's data thus becomes secondary data (Hendrix *et al*

2003). Although the auditors are trained and credentialed to complete the quarterly documentation, their lack of clinical nursing competency for that situation affects the accuracy of the data records (Shin & Scherer 2009). Furthermore, since the auditors have various clinical backgrounds, they are applying different clinical assumptions and analysis when questioning and interpreting the primary carer's observations and records (Lum, Lin & Kane 2005). Thus, the clinical and communication competencies of the auditors are a further threat to the reliability of the collecting process and the validity of the data (Anderson, Buckwalter, Buchanan, Maas, Imhof *et al* 2003).

This whole process is further exacerbated by the inability of inadequately trained NAs to observe and identify the relevant symptoms and signs required (Resnick, Cayo & Pretzer 2009). The RN in charge of a unit may have a global overview of the patient status but lack the detail as they fulfil a managerial role of overseeing the NAs direct care to the patients. The RNs may not have the capacity for direct patient functional observations and recording scores to complete a data chart as they are usually overburdened in managing the patient's nursing needs (Shin & Scherer 2009). As the NAs are the nursing team's primary carers and direct observers the RNs role as manager of the team is to render the in-service training, testing and supervision of NAs who, in the researchers opinion, should record the functional changes as they occur and report it to the RNs. The RNs task is also to render in-service clinical supervision to the NAs (Brunero & Stein-Palbury 2008), a process of professional support and learning in which NAs are assisted in developing their practice of restorative nursing (patient observation, scoring, recording, applying restorative nursing skills) through regular discussion time with experienced and knowledgeable colleagues. The model of recording of RAI-MDS data by credentialed multi-disciplinary external auditors who has no direct care responsibilities towards the patients, and therefore, have to involve NAs who are untrained in observing skills, do not provide accurate data (Shin & Scherer 2009).

Thus, for the sake of accuracy it is advisable for the primary care providers and direct observers to do the recording routinely while they observe changes in

patient functioning, provided they are trained to recognise the symptoms and signs they are dealing with daily. With 90% of the direct restorative care being rendered by NAs ([Resnick et al 2009](#)), they became the most appropriate persons to observe and score the patient functional changes as it occurs ([Schnelle, Wood, Schelle, Simmonds 2001](#); [McCurrin 2002](#); [Hendrix 2003](#)). For this reason the measures have to be designed with the direct observers in mind.

Subjective instruments are at risk of providing biased data. Objective instruments through structured observations of the patient's physical, cognitive and behavioural changes are more likely to be recorded accurately than subjective instruments that require patient recall and interviews that are at a higher risk of discrepancies because recall bias or rater prejudice may influence the data records ([Shin & Scherer 2009](#)). However, primary caregivers can be trained to directly observe and record, using listening skills, the patient's underlying mood and mental changes over time. This is possible if such changes are observed daily in a standardised format and not quarterly.

There are also concerns in applying periodic rather than routine measures. "Periodic" refers to cross sectional, irregular, sporadic, interrupted or recurrent measures while "routine" refers to habitual everyday measures based on methodological measures to observe and score patient changes. All readings on patient functional status are subjected to fluctuations pending subjective judgements, acute episodes, changes in medication schedules, mood alterations ([Fisher, Bergio, Thorn, Allen-Brugge, Gerstle & Roth 2002](#)) and as a result, periodic measures, may not reflect outcomes as accurately as routine measures which may render more stable linear lines of performance over time. Another benefit of routine measure is the evolving decrease in measurement error as refinement through routine clinical supervision of the rating process takes place. This will reduce measurement errors ([Bialocerkowski & Bragge 2008](#)) when the measurement's training material does not provide clear standardised and consistent item definitions that match the raters' observational skills and the clinical connotations attached to each score ([Shin & Scherer 2009](#)). Iterated scores will make clinical and operational sense to both the raters and their patients if done routinely.

Therefore, routine measurements are anticipated to be an ongoing nursing practice with the observations becoming a standard practice. Underpinned by continuous in-service education and clinical supervision the integrity of data will be refined. Training of NA's to contextualise what they observe in patient functionality as vital information is an evolving process to improve quality outcomes. However, according to [Black, Lewis, McIntosh and Callay \(2009\)](#), implementing a routine outcome measurement program is not without problems. If the measures used are observational measures, the focus must be on the primary caregivers to record what they observe, and subjective assessments such as patient recalls should be limited. Moreover, the item definitions should match both the skills levels and the scope of practice of the raters to prevent measurement error. The clinical utility of the measures should reduce nursing burden and increase usefulness to nursing administration and supervision. The nurses-in-charge must accept their role of embedding the process into the nursing unit and be prepared to use patient-centred data instead of nursing process-centred data to improve quality of care. This affects in-service training and clinical supervision to increase data accuracy and outcomes ([Resnick et al 2009](#)).

3.4 Rasch Measurement Model (RMM)

In health care sciences the concept of construct validity refers to how well a scale correlates with the construct that it purports to measure in order to accurately operationalise the concept it measures ([Linacre 2010](#)). Although the RMM has been widely used in the field of education over the last 40 years, this method of validating construct only became popular in the health care sciences over the past decade with the reporting of a variety of health care measures being validated by the RMM ([Tennant & Conaghan 2007](#)). Because of its relatively new application in healthcare measurement validation, the researcher reviewed the evolution of the Rasch model in order to achieve a higher level of comfort with the method of analysis selected for this study.

In the 1920's there were heated discussions amongst physical and social scientists on the question: Is psychological measurement at all possible? At that time, Norman Campbell - a most influential physicist - believed that in order to measure, one should be able to perform a physical operation, a concatenation, such as placing rods end-to-end to measure length or piling bricks one on top of another to measure weight (Campbell 1919). The physicists - who became known as the "hard" scientists - were unequivocal in their response to the social scientists – known as the "soft scientists": *"No, you cannot use measurements, because measurement requires a deliberate action, a concatenation and you cannot concatenate a person's head!"* (Campbell 1921).

The "soft" social scientists had no answer to that, but years later the indirect answer came from the social scientist Stevens (1946), when he published his hierarchical theory for measures opening the door of measures to the social sciences. He advocated the assignment of numbers to objects, events, observations or experiences (qualitatively-ordered scores) according to a set of developmental rules (latent variable); and thereby concluded that some form of measurement exists that is available to all scientists. He described four levels of measurement:

- nominal (classification or grouping),
- ordinal (assigning numerals to represent a ranking or a rule but with no regard to equal spacing between the numbers, thus not useful for adding and subtracting of scores),
- interval (assigning numbers in hierarchical way with equal spacing in between, thus useful for adding and subtracting and totalling of scores), and
- ratio (same as interval, but with an absolute zero point on the scale allowing for multiplication and division).

Since Steven's publication the "soft" scientists started using the nominal and ordinal measures and used whatever numbers they could apply to an observation, provided there is a ranking or a rule attached to such number. This has caused confusion as these qualitatively ordered scores were used as

absolutes (having linear equal spacing between the numbers) and advanced statistical models were developed to analyse the outcomes of the measured observations. However, the “soft” scientists still find it difficult to comply with the “hard” scientists who laid down the measurement definition as: *“measurement means that adding one more unit, adds the same amount extra, no matter how much there already is”*; such as with length the constant would be one more meter and with weight the constant amount would be one more kilogram and in temperature the constant amount would be one more degree (Campbell 1919). However, at that time Campbell also could not imagine how to constitute a constant unit for any human function or attitude on a linear scale, and neither could the social scientists during those years. Therefore they gave up on reforming ordinal scales into interval scales and instead focussed on advancing their statistical analysis of ordinal scores to overcome the lack of measurement integrity.

To date the practice of supporting research findings on qualitatively-ordered data is been heavily criticised. Bond and Fox (2007) stated unequivocally that it is no longer good enough to assign numerals to represent a ranking or a rule to human function or behaviour and presume those numbers have measurement properties or assert them to be data collecting measures of the health sciences. Although both human and social scientists are aware that it is likely beyond their capacity to develop an absolute zero starting point necessary for ratio-level measures, the possibility of improving ordinal scales into interval measures remained the challenge up to today. However during the 1950’s to 1980’s, this challenge became an impasse for the social and healthcare scientists.

In the early 1950’s, unaware of the scientific disputes on measurements, George Rasch, a Danish mathematician (Rasch 1960), resolved the challenge. In trying to find a solution for a particular problem the Danish Department of Defence had with educational tests, Rasch discovered the relationships between human ability versus item difficulty and concluded a logic that was not considered before. The underlying principle Rasch detected in a data matrix of

a well constructed dichotomous test was that *“a person having a greater ability than another person should have the greater probability of solving any item of the type in question, and similarly, one (test) item being more difficult than the other means that for any person the probability of solving the second (test) item is the greater one”* (Rash 1960 p 117). This principle led him to devise a mathematical model to develop rules for a hypothetically perfect fundamental measure for social scientists, today known as the Rasch Measurement Model.

The extent of Georg Rasch's discovery only became apparent much later through Wright (1977) who rediscovered the value of the Rasch model and popularised it for fundamental measurement development in the educational sciences. The perfection of the RMM lies in its simplicity which also renders it applicable to all human sciences and is *“currently the closest generally assessable approximation of fundamental measurement principles in the human sciences* (Bond & Fox 2007 p 14). However, in the mid 1990's the similarities in the education and the restorative health sciences became apparent when posing the key Rasch questions: When a person (student or patient) with this ability encounters an item (test or task), what is the likelihood that this person would get it correct? Since discovering this correlation, the Rasch model has been successfully applied and published in the health sciences in the past decade (Tennant & Conaghan 2007).

Kottorp (2003) advises developers of Rasch measures to start off by conceptualising a variable as a single unidimensional construct represented by a straight line. Secondly, the developer should imagine placing people on this line based on the idea that they each have more or less of the ability that is conceptualised by the line. Thirdly, the developer should design equal stepping stones (or items) on the line from easy to difficult which will determine the range of the test. The sensitivity of the test is then determined by how many steps are placed on the line, how closely they are positioned, and how well they match the ability of the persons in the sample to be tested. The proposed test for validation is then evaluated by gathering the performance data on the sample of people being scored on the latent trait and analyse it.

The Rasch measurement system is modelled in a unique way to establish the property of invariance by ordering the persons according to their level of ability on the straight line (or trait being measured) and ordering the items according to their difficulties. It follows therefore that the person with the higher ability (or more of the trait) should always have a higher probability to get an item correct than a person with a lower ability (or less of the trait), no matter which of the items they encounter. Similarly, a more difficult item should always have a lower probability of being answered correctly than an easier item, regardless of the ability levels of the persons who performs those tasks (Iramaneerat, Smith & Smith 2008). This is also referred to as the Latent Trait Theory, a model based approach, in which latent trait estimates depend on both patients' responses and on the difficulty of the items that were used to obtain those responses (Embretson & Reise 2000). The RMM contributed hugely to the evolution of the Latent Trait Theory which currently addresses most of the many shortcomings of the Classical Test Theory approach to data analysis (Wright & Mok 2004).

The original RMM was invented for dichotomous (yes/no) measures, and the Rasch relationship equation of the simple dichotomous formula was as follows:

$$B_n - D_i = \log (P_{ni} / (1-P_{ni}))$$

where

B_n = Ability measure of person n

D_i = Difficulty calibration measure of item i

P_{ni} = Probability of a correct response from person n on item i

$1-P_{ni}$ = Probability of an incorrect response from person n on item i

In non-mathematical terms: the logarithm of the odds ratio between the probability of passing an item and the probability of failing an item equals the difference between the ability of the person and the difficulty of the item. More explicitly, the Rasch analysis enables the calibration of item difficulty (e.g. where D_i is placed on the straight line) and person ability (e.g. where B_n is

placed on the same straight line). As both these calibrations are expressed in logits (log-odds probability units), they are additive in nature. (Kottorp 2003).

The basic Rasch model provide two specific facets of insights, namely; the easier the item the more likely it is for a person to successfully overcome it; and secondly the more able the person the more likely it will be that they will overcome the more difficult items than those persons with less ability (Wright & Stone 1979). The advanced Rasch models also provide a third facet, namely; rater accuracy or severity. Although this facet accounts for a substantial amount of variance in data (Linacre 1989), it was not considered for this thesis continuing with the two basic facets; imagine the odds of succeeding or failing an item is 50:50. Therefore $B_n = D_i$ because the ability of the person is equal to the difficulty of the item. The logarithm is then calculated to 0 and the odds ratio = 1. But if the person's ability is higher than the item difficulty (e.g. $B_n > D_i$) one would expect to see the probability of succeeding in the item would increase and, conversely, the probability of not succeeding in the item would decrease. This would result in an odds ratio of larger than 1 and a logarithm of the odds ratio larger than 0. On the other hand, if the person's ability is lower than the item difficulty (e.g. $B_n < D_i$), the probability of succeeding in the item would decrease, and the probability of not succeeding the item would increase. This would result in an odds ratio of lower than 1 and a logarithm of the odds ratio lower than 0 (Kottorp 2003).

The original RMM did not take into consideration the polytomous (strongly disagree, disagree, neutral, agree, strongly agree) measures invented by Likert (1932). However, in 1978, Andrich published a conceptual breakthrough article in which he noted that a polytomous Likert rating scale could be thought of as a series of Rasch dichotomies. This enhanced version on the original work of Rasch became known as the Rasch-Andrich Rating Scale Model.

However, over time more complex polytomous social measures came to the fore in which the different items in the same scale each have its own rating scale structure. This development forced the Rasch converts back to the

drawing board and in 1982, Masters published his Rasch-Masters Partial Credit Model which he invented while examining multiple-choice questionnaires and came to the conclusion that some incorrect answers on items are closer to correct than other incorrect answers on other items. He was asking himself how the examinee could receive partial-credit for selecting a partial-correct answer when the “partial-correctness” structure differs from item to item in the same scale. His solution to his own question was that the Partial Credit Model recognises a partial-credit ratings scale as being specific to each item (Linacre 2010).

$$\log_e (P_{nij} / P_{ni(j-1)}) = B_n - D_{ij}$$

The Partial Credit Model, specifies the probability, P_{nij} , that person n of ability B_n is observed in category j of a rating scale specific to item i of difficulty D_i as opposed to the probability $P_{ni(j-1)}$ of being observed in category $(j-1)$ of a rating scale with categories $j = 0$. The rating scale structure (F_{ij}) is now specific to item i . This means that partial credit items with the same number of categories and the same raw marginal scores, taken by the same people, can have different difficulties if the pattern of category usage differs between the items (Masters 1982).

The Rasch fit statistics would reveal how well scales approximate (“fit”) the RMM. If poor fit is achieved poor measurement qualities are reported. However, the RMM guided the researcher along a diagnostic pathway to identify under and over-fitting characteristics in the scales, and if possible provided the remedies to rectify and adjust the weaknesses in the scale structures to a point where a scale could optimally fit RMM. This process of refining scale structure is referred to as scale calibration (Bond & Fox 2007). The degree of final fit to the RMM expectations indicated the level of confidence to which the scale can be used in future as a fundamental measure to produce measurements useful for adding and subtracting and performing parametric analysis (Bond & Fox 2007).

Importantly also, optimal fit to RMM without clinical reasoning and sensibility does not bode good measurement ([Linacre 2010](#)). Therefore, successful measurement estimation in healthcare is reliant on the calibration process to achieve maximum interdependence between two facets; a mathematically perfected Rasch fit and a clinically perfected sensibility. If one of these creates more reassurance than the other, the scale would not achieve success. To attain this balancing act, scale calibration in healthcare should be an on-going process of refinement and the researcher's study may only be the turning of the first stones. Because of this pragmatic approach to scale development, the ALPHA, BETA, GAMMA and DELTA, in their original ordinal format, were subjected to the RMM for fit statistics to establish the baseline level of validity that each scale can achieve. It is useful to establish the baseline degree of confidence whereby ordinal scores can be transformed into linear interval level measurements, and establish their potential for future calibration towards excellence ([Tennant & Conaghan 2007](#)).

Critics of the RMM still argue that one cannot physically align equal bits of human functioning together to produce a fundamental measure same as we align centimetres to add up to a meter. However, one can debate that many measures in the scientific world, such as density, has no fundamental measurement abilities to be demonstrated in concrete units, but is rather derived indirectly through calculating constant ratios between mass and volume. Following on the scientific fundamental measures such as "length" and derived measures such as "density", the RMM for measurement can be considered as a third set of scientific measures mostly supportive of the "soft sciences" ([Bond & Fox 2007](#)).

Importantly, the RMM does not aim to replace the conventional statistics of the social sciences, but rather aims to provide scientific measurement principles to conventional statistics so that the Rasch estimates of patient ability to perform an item with a known difficulty becomes the preferred data for statistical analysis in the human sciences ([Bond & Fox 2007](#)). Having stated the above, several conventional statistical methods using ordinal raw scores are still used to evaluate patient outcomes, which is a major limitation of this kind of analysis as

“raw ordinal scores” are improperly treated as “valid interval data” (Fisher 1993; Bond & Fox 2007; Linacre 2010). Since most statistical parameters (e.g. mean, standard error, and correlation coefficients) are based on interval data, the underlying assumptions of those statistics are violated when applied to ordinal data. Such applications lead to results that mislead clinical interpretations (Fischer 1993; Wright & Linacre 1989; Bond & Fox 2007). In contrast, the RMM is a specific technique, that when applied to well-constructed ordinal rating scales, may transform ordinal scores into linear interval measurements, allowing conventional statistics to analyse valid patient outcomes.

3.4.1 Requirements for RMM analysis

However, not all ordinal data can be imported into the RMM and expected to be transformed into linear interval measures. Certain requirements of the RMM should be met to make valid Rasch inferences from qualitatively ordered data. The most commonly used Rasch models necessitate scale developers and researchers to first consider the four basic scale requirements before the RMM can be applied to re-construct an ordinal scale into an interval scale. These requirements are unidimensionality, local dependency, monotonicity and invariant item ordering. Only when a scale’s raw data approximates the Rasch model on the following four basic requirements, can further measurement construction proceed using the RMM (Iramaneerat, *et al* 2008):

- Unidimensionality assumes that all the items on the scale measure the same underlying trait the scale is intended to measure e.g. measuring the severity of mental functioning as in the DELTA. The DELTA items are expected to contribute different kinds of information about mental functions (underlying latent trait). When two or more items contribute to a different trait that underlies the DELTA, there is a concern about a secondary dimension in the DELTA (Sijtsma & Molenaar 2002; Bond & Fox 2007).

- Local independence assumption is that firstly, data is not collected from the same persons more than once, and secondly, after accounting for the effect of the unidimensional construct of interest there should also be no significant residual correlation between the items. For example, if one cannot perform the item: “dress lower body” due to limitations to lower body function, that person can also not do the item: “toileting” as both items require the same function of pulling up and down pants. The response of a person to one item should not be influenced by his or her response to any one of the other items. Therefore, if after accounting for the effect of lower body function already, there still exists significant correlation between the two items, that is local dependence (Smith 2005).
- Monotonicity means that the probability that a person will respond to the categories of the items is monotonically non-decreasing over the range of the latent trait. In other words; a severely ill patient with low latent trait ability should always have a lower probability of responding to a DELTA item category than a moderately ill patient with more of the latent trait ability; regardless which item on the DELTA encountered. Similarly, a more difficult category (e.g. score 6) should always have a lower probability of being responded to correctly than an easier category (e.g. score 2), regardless of the mental ability of the patient (Sijtsma & Molenaar 2002).
- Invariance of item response functioning is a characteristic of an Item Response Function curve that showed a constant slope variable, making all the curves not intersecting with one another (Bond & Fox 2007).

3.4.2 Measurement expectations from the RMM

Over and above strengthening and validating the above four measurement requirements, the RMM is further useful to reconstruct and improve scales with desirable measurement properties such as:

- Creating a singular measure. By placing both the persons' ability and items' difficulty facets on the same scale to allow direct measurement comparison between them. In order to do this, the RMM puts both facets on a common logit scale. The logit scale is the logarithm of odds of the probability of achieving a correct response over an incorrect response ([Iramaneerat et al 2008](#)).
- Creating a linear interval measure. As the RMM is a form of additive conjoint measurement, the new scale is predicted to function as an interval measure when the scale fits the RMM requirements satisfactorily. ([Linacre & Wright 1989](#)).
- Creating an objective measure. The RMM deliver on the basic requirement of objective measurement by separating the parameter estimates. In other words, when the data fit the model expectations, the measures of patient ability remains the same regardless of which subset of items is used for the fit analysis ([Stone 2004](#)).
- Creating reliability. High accuracy in calculating internal consistency reliability is achieved, as each of the individual item difficulties and person ability measures has their own unique standard error of estimate ([Smith 2004](#)).

3.4.3 Fit statistic indices of the RMM

The inferences from the RMM analysis on the construct validity of the scales are made from the degree of fit achieved. The RMM model provides a range of indices to monitor how the quality of the scales' fit to the RMM changes during the calibrating process. These calibrating indices are referred to as the fit statistics ([Wright & Masters 1982](#)). According to [Linacre \(2010\)](#), the most commonly used RMM criteria and indices for fit statistics include the following:

- Outfit and infit statistics are reported as Mean Square (MNSQ) statistics. Both outfit and infit values has an expectation of 1.0 and

can range from 0 to infinity. Generally, outfit values larger than 1.0 indicate more variability than expected in the data. Conversely, outfit values less than 1.0 indicate less variability than expected in the data. Whilst outfit stats are more sensitive to anomalous responses by over or underperforming persons, particularly in tests with wide ranging latent variables; the infit stats informs anomalous responses by average persons near the centre of the latent variable distribution. Values greater than 2.0 are of a great concern; as such a value indicates the scale is inaccurate [Linacre \(2010\)](#).

- Point measure correlation (PT MSE CORR) should report a noticeably positive correlation of >0.3 which will confirm that the distribution and direction from easy to difficult on the latent variables is aligned with the severity of the patients. RMM expects the lowest category on the latent variable to be easier for severely ill patients than the highest category [Linacre \(2010\)](#).
- In RMM, the reliability is an indication of reproducibility and it is reported in both person and item reliability. Person reliability values below 0.8 indicates that the number of items in the test are not enough to represent the latent trait; and item values below 0.8 might indicate that the sample used for analysis is too small to test the latent trait [Linacre \(2010\)](#).
- The raw score variances explained by measure is reported in both empirical and model values. The empirical variance components are for the observed data, and model variance is the expected components when the data fit the Rasch model well. When these two values are both high values and match reasonably, another noticeable indicator of fit to the Rasch model has been achieved [Linacre \(2010\)](#).
- Rasch category statistics indicate how satisfactorily the categories of the items are working for the sample. This level of analysis presents the researcher with category probability curves to enable investigation of category functioning. **Figure 3.1** shows the ideal fit where the peaks of the categories are all in ascending order along

the latent variable, and each category in turn is the more probable of any other one of the categories. Furthermore, the cross-over points between the categories are ordered; e.g. the descending curve of each category clearly crosses the ascending curve of the neighbouring category. These cross-over markers are the equal probability points or the thresholds or the parameters of the Partial Credit Model (PCM). Such ordering of the categories indicates a good fit to the RMM. Category statistics that conclude a disordering of categories require remedial action which can be provided by the WINSTEP software [Linacre \(2010\)](#).

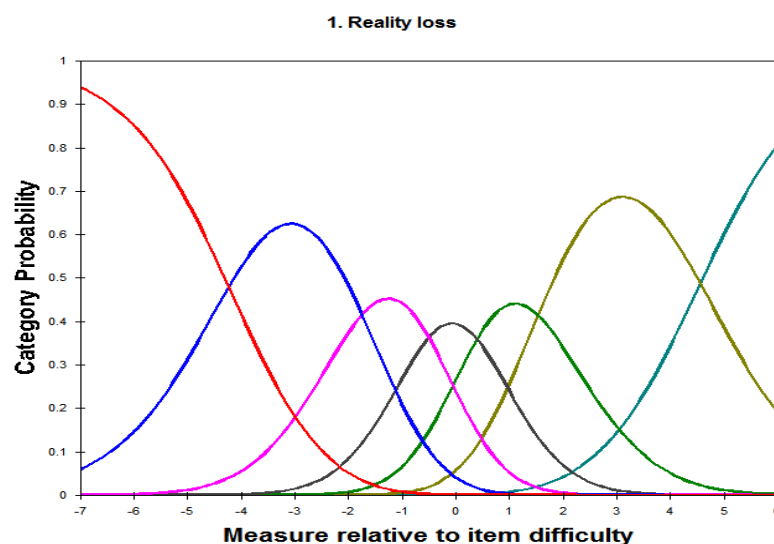


Figure 3.1: The Category Probability Curve of the first DELTA item 1: Reality Loss.

3.4.4 Category analyses guidelines of the RMM

When the above fit statistics indices of a scale are available, the inferences of the scales potential to satisfy the RMM are better understood. However, the four nursing measures under investigation are all polytomous scales and as such additional attention must be given to their rating scale structure by investigating how the scales' categories are functioning. [Linacre \(2004\)](#) proposed a checklist with the following eight guidelines to analyse category functioning:

- Are all the items orientated with the latent trait?
- Are there at least 10 observation frequencies in each rating category?
- Are the observations regularly distributed across all rating categories?
- Are the average measures advancing monotonically with the category?
- Are the outfit mean square values less than 2.0?
- Are the thresholds advancing with the categories?
- Are the thresholds advancing by at least 1.0 logits for a five-category or more rating scale?
- Are the thresholds advancing by not more than 5.0 logits?

3.5 Summary

In this chapter the literature review on three phenomena relevant to this study have been discussed. Firstly, restorative nursing, being a link between nursing care and therapy and largely unknown in the South African nursing practice, have been reviewed. The focus was on its need for routine nursing measures to be used by primary caregivers being in the best position to measure patient function through direct observations. Secondly, nursing utility was explored and confirmed as an integral facet to the validation of a nursing measure which is required to provide routine data from the nursing process. Finally, the property of the Rasch measurement model was explored to be introduced as a instrument to analyse the construct validity of the four nursing measures under investigation. Further than construct validity, the added value of the RMM to also calibrate qualitatively-ordered ordinal scores, through probabilistic inferences, into quantitatively-ordered linear interval measurements have also been discussed. In Chapter Four the researcher will discuss the methods used to address the research problem of analysing and reporting scientifically on the four measures of nursing utility and construct validity.

CHAPTER FOUR: THE METHODS

"All measures are numbers ... but not all numbers are measures."

Ben Wright 1997

4.1 Introduction

The purpose of this study was to make available validated measures to the restorative nursing process. For this purpose the researchers has developed four nursing measures which he assumes has both significant nursing utility and construct validity. This construct theory of the researcher created the research problem of the study, which is to find a scientific method to examine and validate the researcher's expectations.

However, in Chapter Four, the research problem of the study was resolved when a scientific research method was tabled to test the researcher's construct theory of validity of the two research objectives of the study. A qualitative method was proposed to investigate the nursing utility and a quantitative method was designed to analyse the construct validity of each of the ALPHA, BETA, GAMMA and DELTA scales. Jointly and equally, these qualitative and quantitative methods presented in Chapter four satisfied the research problem and research objectives of the study.

4.2 Design for the entire study

Both a qualitative and a quantitative design were used in the study. The qualitative approach investigated the nursing utility properties while quantitative techniques investigated the construct validity properties. Although both these

qualities jointly and equally contribute to the validity reporting of the measures, their enquiries have different designs, study populations, sampling, settings, methods of data collection, methods to ensure data quality and methods of data analysis. For this reason the two methods are reported on separately in this chapter, but discussed jointly as an interdependent characteristic in Chapter nine: Conclusions.

Table 4.1: Overall research design of the study.

| Validation Study | | | | |
|---------------------------------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|------------------------------------------------------|------------------------------------------|
| First Study: Nursing Utility | | | Second study: Construct validity | |
| Qualitative Design | | | Quantitative Design | |
| <i>Strategy: Non-Experimental/ Descriptive/ Cross Sectional</i> | | | <i>Strategy: Methodological/ Cross Sectional</i> | |
| Nursing process | Uniform language | Quality assurance | Validity | Reliability |
| <i>Focus group of nurses; descriptive data</i> | <i>Focus group of nurses; descriptive data</i> | <i>Focus group of nurses; descriptive data</i> | <i>Patient functional raw score data</i> | <i>Patient functional raw score data</i> |
| <i>Non-Probability purposive & convenience sampling</i> | <i>Non-Probability purposive & convenience sampling</i> | <i>Non-Probability purposive & convenience sampling</i> | <i>Non-probability quota sampling</i> | <i>Non-probability quota sampling</i> |
| <i>Deducted content analysis</i> | <i>Deducted content analysis</i> | <i>Deducted content analysis</i> | <i>Rasch Analysis</i> | <i>Rasch Analysis</i> |

4.3 First study: Nursing utility

The nursing utility study verifies and reports the degree of usefulness the measures provided to the nursing profession when implemented routinely within the nursing process. Broadly, it includes proficiencies to establish and address

patient needs, restorative nursing needs and the rendering of patient based information.

4.3.1 Design of the nursing utility study

The original intention was to explore the nursing perception of nursing utility using a quantitative design with questionnaires. However, as the study progressed, the reality of limited numbers of nursing respondents (<10) available to complete a standardised questionnaire for every individual nursing measure became a reality. These small numbers would have resulted in datasets too small for meaningful statistical analysis. The design was thus changed from a quantitative investigation using a questionnaire, to a qualitative study using focus groups to explore the nursing utility phenomenon.

[McLafferty's \(2004\)](#) cautioned about the difficulty to recruit a sizeable number of nurses at any one point in time to participate as respondents in research. According to her, this unavailability of nurses to participate in research is due to the nursing shift system that includes day / night duties, days off schedules, annual leave calendars, and general absenteeism. [Webb \(2002\)](#) also advised to take into consideration high levels of non-attendance when designing research studies. Taking this into consideration, the nursing managers anticipated that 3-6 nurses can be made available for a one hour focus group at session in a facility. This small number of experienced nurses is adequate for meaningful qualitative data collection purposes provided they have experience in the phenomenon under investigation to provide rich and quality data ([Carey 1994](#); [Kreuger 1994](#); [Morgan 1996](#); [Kritzinger 1996](#); [Twinn 1998](#)). Hence, for practical reasons to overcome the relatively small numbers of nursing respondents available, focus groups were accepted for qualitative data collection whereby nurses would render evidence on the nursing utility as experienced in their facility after implementing the measures for more than three months.

Following on the decision to use focus groups, the second consideration was what technique of data analysis to be used as this would impact on the researcher's approach to the focus groups. This required taking a step back and reflecting on the design and development of the measures to be investigated. These measures were a product drawn from the intuitive knowledge of practising nurses, who collaborated with the intention to produce measures that comply with specifications of nursing utility. The product thus had a high probability of nursing utility, as the collaborating nurses designed it to suit their own working environment. As a result, the nursing measures should achieve high content validity on the subject domain of nursing utility. According to [Foxcroft, Patterson, Le Roux and Herbst \(2004\)](#), the use of a panel of nursing experts to be involved in the design, development and review of nursing utility specifications, significantly contributes to the content validity of the instruments. In the researcher's study the expert nurses would have reviewed the concepts contributing to the utility domain.

Thus, with an existing theory and knowledge in place, the purpose of the design was to test an assumption of nursing utility in a different environment. The knowledge and understanding of the expert nurses was tested in the real nursing environment. Therefore the researcher had to create scientific structures of investigation, data collection and analysis on the basis of previous knowledge.

[Elo & Kyngäs \(2007\)](#) reported that the need to validate an existing theory is commonly experienced in nursing studies. For this purpose they recommended the qualitative approach of deductive content analysis, which is particularly useful when the general objective is to test a previous theory in a different situation or to compare categories at different time periods, or testing concepts, models or theories. For this reason, a deductive content analysis design was adopted to verify the existing clinical utility knowledge of the measures. This approach dictated both the design of the question guidelines as well as the analysis of the data.

4.3.2 Population, sampling and settings

The purpose of the focus groups was to explore the nurses' experience when implementing the nursing measures into their restorative nursing practices. This exploration required investigations into the nurses' perception of the degree of usefulness the measures add to their daily practice. At the same time the study would be interested to gather information on any further recommendations or concerns the nurses may have, or any oversights in the design of these measures they might have experienced.

Therefore, the study population consisted of the appropriate nurses, nursing assistants or caregivers, who were trained, tested and credentialed in the use of one of the four nursing measures and having worked with the specific measure in the clinical settings for at least three months or more. Credentialing meant they received a training manual, undergone a full day training, did a one week practical on scoring their patients under supervision, followed by a written open-book four-case-study test on which they have achieved a minimum of 80% pass-rate. Credentialed nurses were considered to have extensive experience of implementing and applying the measures on their patients with functional deficits in sub- and non-acute nursing settings, and having recorded scores allocated to their patients. They have also reflected on the meaning of the scores allocated and have considered possible options to improve on the scores.

The sampling thus followed a purposive sampling approach whereby credentialed nurses have been invited to participate in the focus groups. Invitations were extended three weeks in advance via the nursing manager to credentialed nurses being available and willing to participate as respondents on a scheduled date and time selected by the nursing manager. It was thus also a convenience sample. No exclusions based on gender, age, race, social groupings or religion were imposed and no credentialed nurses were excluded.

The settings of the focus groups were selected to comply with the criteria set by [Dilorio, Hockenberry-Easton, Maibach and Riverio \(1995\)](#) as being familiar, neutral and non-intimidating to the participants. The nursing managers were required to select a venue in their facility that complied with these mentioned criteria for the specific number of participants.

4.3.3 Data collection

Only one focus group provided the qualitative information at a facility. If the measure under investigation, as in the case of the BETA, require investigations into managerial data provided by professional registered nurses (PRNs), and observational data provided by nursing assistants (NAs) or caregivers(caregivers); two separate focus groups were established; one with the PRNs or one with NAs or caregivers.

At the onset of each focus group, the researcher explained the scientific processes to be followed; the confidentiality to be adhered to and provided the ethical clearances obtained and the resultant guidelines to be observed to all of the respondents involved in the focus group. After all questions were asked and answered, the voice recorder was activated, indicating the start of the formal data collecting process.

The researcher entered the data process by tabling the first primary question from a structured interview guide, which consists of primary and secondary questions ([Table 4.2](#)). Although these questions were designed to verify the existing theory that the measures have a nursing utility, they also served as a guide to elicit as much data as possible. Therefore, whenever the researcher felt the need to stimulate discussion or further clarity, he freely asked any other questions or use comments as seemed appropriately necessary ([Elo and Kyngäs 2007](#)). Participants were encouraged to freely enter into a general debate on the usefulness of the specific measure, and space was also created

to identify, table and discuss ambiguous matters that may arise. The focus group sessions were audio-recorded and transcripts were made.

Table 4.2: Structured questions for the focus groups.

| No | Primary questions | Secondary questions |
|----|-------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | <i>Are the measures useful to the nurses?</i> | Adequate training and testing; scores adequately reflecting patient function and changes; scores supporting the nursing assessment, diagnosis, restorative techniques and evaluations. |
| 2 | <i>Can it be embedded into the nursing process?</i> | Time, effort, difficulty levels, routine recording in patient file, level of consensus and agreement, daily done. |
| 3 | <i>Can the scores be used as a uniform language?</i> | Use in communication with patient, families, amongst nurses, between multidisciplinary team members, funders etc. |
| 4 | <i>Do the measures improve the quality of nursing?</i> | Improve nursing skills, increased awareness of patient functional needs and patient based outcomes, setting patient goals, objectives and benchmarks. |
| 5 | <i>What are the perceived problems with the nursing measures?</i> | |

The researcher put special emphasis on the search for any “weak links” in the measures that might violate nursing utility (e.g. referencing inappropriate nursing techniques in a particular scope of practice). Focus groups were particularly helpful when consensus in the group was able to verify the very existence and extent of such experiences. Also helpful was the approach to elicit solutions from the participants on how to overcome any future threats to nursing utility. Threats would allow the developer the opportunity to revisit, rectify and enhance usefulness of the nursing measure.

4.3.4 Data quality

Several measures, over and above a commitment to rigorous collection of high quality data and honest reporting, were taken from the literature to generate valid information important to the advancement of the nursing utility of the measures. An example being the practical advice from [McDaniel & Bach \(1996\)](#) to first serve quality refreshments before the start of the formal scientific

procedures to create a relaxed atmosphere and put the respondents at ease, allowing them not to withhold rich information. This informal atmosphere was subtly used to ease the group into a more formal structure when the ground rules were explained to the participants. The use of the voice-recording was explained as well as the transcript to be made afterwards.

Sampling was used to increase data quality. Professional registered nurses and nursing assistants were not mixed in the same focus group. The reason for separate groups was that the research objectives to be explored with the professional nurses were the higher level managerial issues of implementing the nursing measures, whilst the research objectives with the nursing assistants were more directed to issues pertaining to the ease of observing and scoring the patients. However, [Carey \(1994\)](#), [Krueger & Caseyger \(2000\)](#), and [McLafferty \(2004\)](#) validate this decision, but for another reason. According to them the value of inviting respondents with similar characteristics to a single group; e.g. a homogeneous grouping, increase the quality of the data as it allows for specificity and relevancy and therefore have a low risk of having to do multiple sessions to verify the data collected.

To add further strength to the notion that quality data are dependent on the homogeneity of the participants in the group; the researcher stipulated that the participants in a focus group should be familiar with each other. This was achieved as the focus groups were accommodated in the facility where the respondents were working together and where they were all familiar with the patients from whom they drew their inferences. This arrangement was initially a pragmatic decision, but the quality benefit from homogeneity soon became apparent as the participants could immediately contextualise their views within the group and verify their viewpoints with concrete examples.

[Morgan \(1997\)](#) has tabled a questioning methodology useful to maintain a good conversation free from both bias and generalisation. The researcher was advised to keep the discussion focussed by skilfully identifying the useful

generalisations in the discussions and then use structured questions to guide the generalisations into those very specific issues that will deliver the required data. The structured broad based questions in Table 4.2 were designed to start-off with general discussions, allowing the researcher with rewarding opportunities to guide discussions into those specific issues that will provide rich data on nursing utility.

Krueger and Casey (2000) suggest researchers could increase the quality of the data if they make a concerted effort to create a naturalistic environment. By achieving that, the researcher not only achieves content data but is also be able to explore the meaning and depth behind it, as presented in the form of emotions, contradictions, irony and tensions. For this reason participants were informed that any strengths, weakness, opportunities or threats identified would be considered a positive contribution and that any suggestions, ideas or thoughts put to the researcher to improve clinical utility, would be appreciated. The researcher ensured that every respondent fully understood that all comments, discussions or observations were useful.

4.3.5 Data analysis

Every step during the design and development of each measures was tested with the question: would this change or addition improve nursing usefulness and to what extent? Expert nursing respondents, with extensive practical experience, were entrusted to validate these questions. They were given four categories (Table 4.2) of clinical utility to consider, namely: overall usefulness; ease to embed in the nursing process as a routine measure; ability to become a universal language as measure of severity of illness and potential to improve quality of nursing care. With these categories been tested, the measures had nursing utility, and the challenge to the researcher was to re-test existing knowledge and understanding in another but similar facility.

For this reason, a deductive (also referred to as “directed”) approach to content analysis was selected to re-test the theoretical assumption that each measure has existing nursing utility (Elo and Kyngäs 2007). As the goal of a deductive content analysis is to validate an existing assumption, it provided the researcher with a focused approach to the enquiry of nursing utility. As some prediction already exists about the variables of interest, these variables provides a conceptual framework of analysis to predetermine an initial coding scheme according to the assumed categories (Hsieh & Shannon, 2005). The same four categories used by the original nursing collaborators to anchor the design for nursing utility were used in the researchers study to validate the existing knowledge. In order to simplify matters, the same four broad questions (categories) that guided the developers towards nursing utility, were again posed to the focus groups to test their experiences on the same matter. Furthermore, the same four questions became the four categories in the matrix of deductive content analysis. A fifth category was included to explore unexpected considerations that might require remedial work.

A preliminary coding of qualitative data was established by listing the positive quotes and the negative quotes under each of the four categories. For each quote listed the researcher rendered an analysis by contextualising the validity of the quote. Similar subject matter and themes were grouped or collapsed if needed. The quotes with their analysis were prioritised in the degrees of acceptance or rejection of clinical utility. This was followed by a synthesis of the participant’s experiences by tabling the positive and negative points and the recommendations to rectify the negative characteristics. Finally, an overview discussion was rendered on the outcomes of the clinical utility of each measure.

As the reporting of each scale was done independently in Chapters Five, Six, Seven and Eight, where the formats had to comply with the page limitations of this thesis, care was taken not to compress the qualitative data close to a point where the integrity of the material might become lost. The rich supporting excerpts and original narrative details were thus preserved in an attempt to

balance the formal data presentation with the original narrative evidence, indicating the depth and context of the data (Polit & Beck 2004).

4.4 Second study: Construct validity

A study into the construct validity of a rating scale should, according to teachings of Linacre (2010), investigate and report the following three basic questions:

- does the scale measure what it intends to measure (Linacre # 2: 31),
- does the scale measure it with precision and accuracy (Linacre # 3:76),
- does the scale provide linear interval data fit for parametric analysis (Linacre # 3: 76).

As the Rasch Measurement Model (RMM) is well suited to investigate these scale attributes, it was used in this study to render testimony of the four scales potential to achieve these basic characteristics. Although the RMM has the ability to investigate more facets of measurement ability, this study is only interested in its person ability and item difficulty facets to establish whether the nursing measures satisfy the basic RMM fit statistics. If achievable, it would conclude a potential for further RMM investigative calibration at a later stage.

4.4.1 Design of the construct validity study

As set out in Table 4.1, the probing into the internal structure of the measures used has a cross-sectional methodological approach. Nursing teams were trained, tested and credentialed to apply the measures through direct observation to score and record patient functionality according to the latent traits defined by the scale structure. The functional scores allocated to the patients were imported into an Excel spread sheet and prepared for the RMM analyses into each scale's construct validity.

The RMM was used for investigating measures because of its strict mathematical expression of the theoretical relation that would hold between the difficulties of measure item and the abilities of patients along a single underlying latent trait. Although no item or patient will ever fit the RMM perfectly, one is however interested in identifying those items and patients whose patterns of responses deviate more than expected. The RMM fit statistics provided these answers. Where required, remedial work to outfitting responses were undertaken to calibrate each measure into maximum fit. The RMM assumes that the raw score data provided by the four nursing measures represented qualitatively-ordered ordinal observations on the intended latent variable. Based on this assumption the RMM measures were computed ([Linacre 2010](#)).

4.4.2 Population

The qualitatively-ordered ordinal data were collected with the four nursing scales under investigation on patients with unpredictable outcomes and in need of restorative nursing care. These patients' outcomes were evaluated in terms of their functional gains achieved. The broad distinguishing characteristics of these patients were:

- the permanency of their functional loss (e.g. spinal injuries versus post-operative recovery);
- the decline or incline of the functional changes (e.g. convalescent versus palliative life situations);
- the modality of functional change (e.g. executive versus cognitive changes; or self-care versus independent living), and
- the speed of change anticipated (normal slow geriatric decline versus rapid end-of-life decline due to terminal cancer).

4.4.3 Settings

The settings and services used in this study included:

- Rehabilitation settings where patients had low probability of complete functional gains e.g. patients with strokes, spinal and/ or brain injuries;
- convalescence settings where patients had a high probability of complete functional gain e.g. complex medical cases or post surgery;
- geriatric settings where patients required monitoring and managing of subtle functional decline;
- mental healthcare settings where patients were admitted with a low executive functioning levels and their functional gains were managed and monitored until discharge.

Although the restorative nursing focus and management models changed across these settings, the basic patient needs to regain, maintain or support human functionality remained the same. The applicable Alpha, Beta, Gamma and Delta nursing scales were used in each of these settings.

4.4.4 Sampling

As the study called for the collection of qualitatively-ordered ordinal patient scores collected by nurses trained and tested in the use of the measures, a non-probability quota sampling technique was followed. The purpose was to draw a sample from the patient population that had similar functional characteristics as the entire population receiving restorative care. This sample procedure relied on the convenience of choice. The aim of the sampling was to replicate observations of personal abilities versus item difficulty along the latent trait presented by the scales ([Brink 2008](#)).

The sample population was not subjected to any further strata or quota, however, strata variables were identified and recorded in the data set to assist in the subsequent statistic analysis and all patients admitted into the settings were included in the sample during the duration of this study.

4.4.5 Data collection

The Alpha, Beta, Gamma and Delta measures were used to collect the ordinal observational scores on patient functioning. The observations were done by the primary caregiving nurses, best placed to do the direct observations of how patients were able to complete the tasks. By applying the algorithms in the nursing measures they converted their observations into patient scores, thus enabling them to arrive and record at an objective number representing the patient's functional status. The scores were collected purposefully under precisely defined conditions in a systematic and objective manner with careful record keeping. The observations were conducted routinely, meaning that at a minimum, the admission, intermediate and discharge scores were completed. The nurse observers were trained to first attempt prompting and cuing before attempting to render functional support to the patients. They were instructed to record only the actual performance of patient abilities as it occurs regularly under their supervision. The potential performance was not recorded by the nursing observers as it occurs during therapy sessions.

As human function is dynamic and fluid depending on influences such as patient emotions, prejudices, values and numerous external socio-economic factors, the observers were instructed to not only observe and record patient scores on admission and discharge, but also to observe and record daily changes in scores as these changes occurred. All the functional score data collected were objective observational data. No subjective, remote, patient self-reporting or proxy information describing the patient functional status were collected.

The admission, interim and discharge patient scores together with identifying stratum such as age, gender, impairment groups, nursing settings were recorded on hardcopy and imported into electronic spread sheets for analysis.

4.4.6 Data quality

This study was reliant on the collection of raw qualitatively-ordered data for mathematical analysis by the RMM. Before the scales were introduced in the nursing settings, pilot studies were conducted whereby the proposed end-users (e.g. nursing assistants) were subjected to the scales for a limited period of testing. This testing period was aimed to identify systematic errors in the decision tree leading towards consistently incorrect scores affecting the reliability of the measures. This included unpredictable random errors due to ubiquitous terms used in defining the items and categories, or situations not substantially relevant to the scorer's observation or environment. After error identification, changes were summarily introduced in an updated version of the training manuals.

Subsequent to these quality adjustments by the end-users, learner manuals and testing materials were drafted and formal face-to-face training sessions were held for primary caregivers and their PRN supervisors. Credentialing was rendered to a facility when the candidates successfully completed 4 - 8 hours of training (depending on the nursing measure), and achieving 80% or more in an open-book test consisting of case studies. Unsuccessful candidates had to repeat the course. Only credentialed nurses may collect data. A record is kept of all the credentialed nurses in each facility.

4.4.7 Data analysis

According to [Bond and Fox \(2007 p 14\)](#), the RMM is "*currently the closest generally accessible approximation of fundamental measurement principles for human sciences*". The extent to which the raw ordinal data fitted the RMM implies the extent of the construct validity of the instrument. The basic Rasch assumption is that (a) each patient is characterised by ability and (b) each item by a difficulty which (c) can be expressed by numbers along a straight line. From the difference (d) between the numbers (and nothing else), the probability of observing any particular score response can be computed.

In this study, the Rasch-Masters Partial Credit Model was applied in the analysis of all four nursing measures. The WINSTEP® Software version 3.70.1.1 (2010) (Annexure G) was used to perform this analysis. A licence to utilise the software was procured through www.WINSTEPS.COM (Winsteps 2010a). To obtain the analytic skills, the researcher participated during August / September 2010 in an online Rasch analysis course “*Practical Rasch measurement – core topics*” sponsored by Statistics.com’s (Annexure H) [www.statistics.com/rasch]. The course leader was Dr Mike Linacre (Certificate 2010).

As the four rating scales were all polytomous measures, where each item was defining its own rating scale structure, the Rasch-Masters Partial Credit Model (PCM) was used to calibrate the four scales. The PCM allowed each scale’s item to define its own category probability structure, allowing one to observe: “Category 1 of Item 3 functions this way” (Linacre 2010 #3).

To analyse the validity of a rating scale, WINSTEPS assumed that the imported scores were qualitatively–ordered observations on a latent trait (items) as prescribed by the developers construct theory. Based on this assumption, Rasch analysed the data to test the dependability of the researchers construct theory which is reported on the variable map, also known as the item-person map, as seen on Figure 4.1. It shows how well the developer managed to arrange the person ability and item difficulty on a common logit scale represented as a straight vertical line. The mean item difficulty is set at 0 logits. The variable is laid out along this line with the most able persons and the most difficult items at the top. On the left, persons are represented by an “#” or “.” (Its value references are reported at the bottom of Figure 4.1). The persons are positioned according to their mean abilities, and on the right, the item labels are positioned according to its mean difficulty. The distribution of the person ability should be matched by the distribution of the item difficulty when norm reference interpretation is required (Linacre 2010).

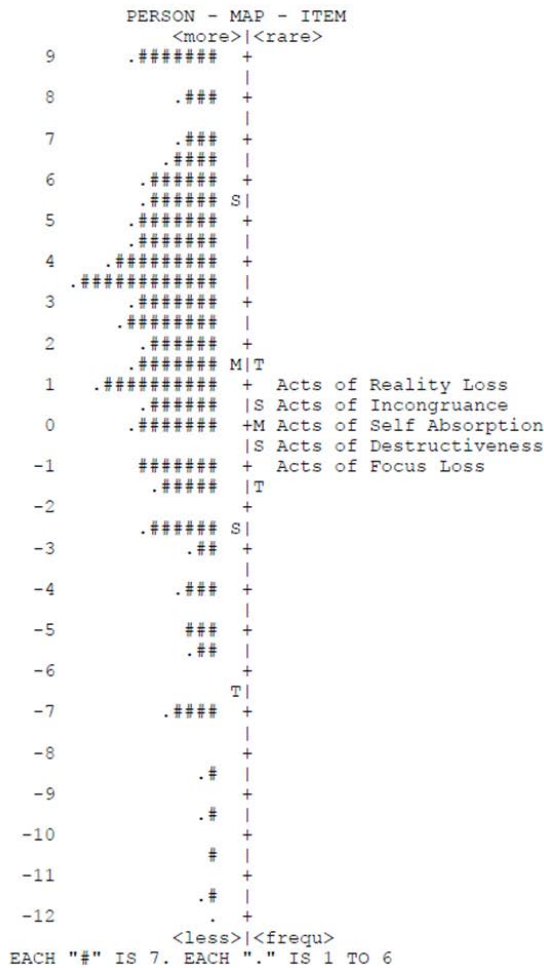


Figure 4.1: Variable map of patient ability and item difficulty in the DELTA.

When examining Figure 4.1, Self-Absorption is the DELTA item calculated to occupy the exact mean difficulty estimate location on 0 logits. This means that a mental healthcare patient with an ability estimate of 0 logits has a 50% probability of succeeding on the Self Absorption item. Furthermore, that same patient would have a greater than 50% probability of succeeding on an item less difficult such as Focus Loss, and inversely, a less than 50% probability of succeeding on an item more difficult such as Reality Loss. With these basic principles evident from the DELTA's variables or item-person map, one can immediately establish that the DELTA might be too easy for the test sample as the patient distribution is top-heavy in comparison to the item distribution. Also

at the top of [Figure 4.1](#) a cluster of scores are observed, meaning that there were data in the sample above the range of the measure e.g. a ceiling effect. A ceiling or floor effect indicates the level above or below where the independent variable was no longer measured or estimated. Clinical or pragmatic evidence should be rendered why ceiling or floor responses are present in the data. In the case of the DELTA, such an explanation was offered in Chapter Eight.

The following valuable stepping stones, provided by [Smith, Linacre, & Smith \(2011\)](#), were followed to analyse the raw data with the RMM:

- Explore the integrity of the data, e.g. missing, folded or nested data and resolve, explain or accommodate the results.
- Examine the map of linear variable as defined by the items.
- Study the map of distribution of sample on linear variables.
- Observe the functioning of the categories and explore the procedures and techniques to improve the fit statistics of the scales such as category collapsing. As this study is not troubled by small sample sizes, informed decisions about the use of categories in the partial credit model should be possible.
- Investigate for secondary dimensions in items, persons, etc, using for example fit statistics and other analysis of residuals.
- Explore Rasch separation and reliability of the measures

4.5 Rasch calibration

The Rasch calibration process of conceptualising, designing, analysing and reporting across the numerous variances involved in an evaluation of a new rating scale is a daunting task. Fortunately for novices, the Rasch experts provided numerous operational guidelines, of which [Linacre's \(2004\)](#) eight basic guidelines to authenticate how well a polytomous rating scale categories fits the RMM is the most comprehensive and logical to follow. Therefore, these eight

guidelines was used as a tick list during the calibration process of category functioning. The guidelines are as follows:

1. All items must be orientated to the latent trait. Winsteps provide an item analysis report where reverse item polarity can be identified.
2. There must be at least 10 observations in each rating category
3. The observations must be regularly distributed across all rating categories.
4. The average measures must advance monotonically with category.
5. Outfit MNSQ values must be less than 2.0.
6. Thresholds must advance orderly with categories.
7. Thresholds must advance by at least 1.0 logits for a rating scale with 5 or more categories.
8. Thresholds must not advance by more than 5 logits.

However, as mentioned in Chapter Three, not any rating scales' raw scores are eligible for the Rasch analysis. Thus, a screening test had to be done to verify if the four nursing scales' raw data did comply with the four basic assumptions for acceptance. These assumptions include local dependency of the data, unidimensionality of the items, and monotonicity of the latent trait and invariance of the structure ([Iramaneerat, et al 2008](#)). As some of these four basic assumption criteria are collaborating with the Linacre requirements, they will provisionally be checked as supportive of the relative Linacre guideline.

- *Local dependency*: The GAMMA dataset used cross sectional raw score data, but the raw dataset of the ALPHA, BETA and DELTA includes the admission, intermediate and discharge longitudinal data. While it is sensible to include all longitudinal scores as it represents the full range of the latent variables, local dependency become a concern when scores of the same patient on admission, intermediate and discharge dates are included. To overcome local dependency, a random sample was created by using an Excel random number generator. This function assigns a random number within a specified range using a uniform distribution to each score entry. A random sample from each of the admission, discharge and intermediate group scores was created. Any duplicate

patients were also then eliminated as far as possible by again running these patients through the Excel random number generator. Thus the residual dataset loaded in Winsteps was local dependency free and all further analysis was done on this random sample (Linacre 2010).

- *Unidimensionality*: If all the items on the scale measure are affected by the same latent trait then there should be only one underlying dimension (domain) present. To calibrate this phenomenon, Winsteps decomposes the unexplained variances in the residuals and report it in five contrasts. The relative fit parameter for unidimensionality is the Eigen value in the first contrast that should be < 2 (Linacre 2010).
- *Monotonicity*: The probability of observing a patient ability (score) must not fluctuate up and down the latent trait but rather follow a monotonic stepwise increase or decrease according to the patient abilities and the item difficulties (Sijtsma & Molenaar 2002). Winsteps analyses an observed average logit measure and produces it in the OBSVD AVRGE column as a rating scale functioning report. It revealed that the scale's observed average measure (logit) behaved as it is suppose to do: it advanced orderly with rating scale categories. Thus, monotonicity was achieved if the OBSVD AVRGE column revealed no disordering of the scale categories. This information is useful to Linacre's fourth guideline.
- *Invariance*: Winsteps produce graphs with Category Probability Curves (CPCs) for each item to view the probabilities between patient abilities and item difficulties (Figure 4.2). The peaks of the CPCs should appear as a range of hills with distinct peaks and clear crossover points between the curve for one category and the curve for its neighbouring category. If these characteristics are apparent, it would provide strong evidence that the scale is complying with Linacre's 6, 7, and 8 guidelines (Linacre 2010).

When a scale conforms to these four basic entry assumptions for Rasch analysis, one can proceed with a higher level of comfort that the calibration of the scale into an interval measure is a probability. However, calibration process cannot be operationalised without first selecting the relevant Rasch indices (parameters or indicators) to track the process of measurement calibration.

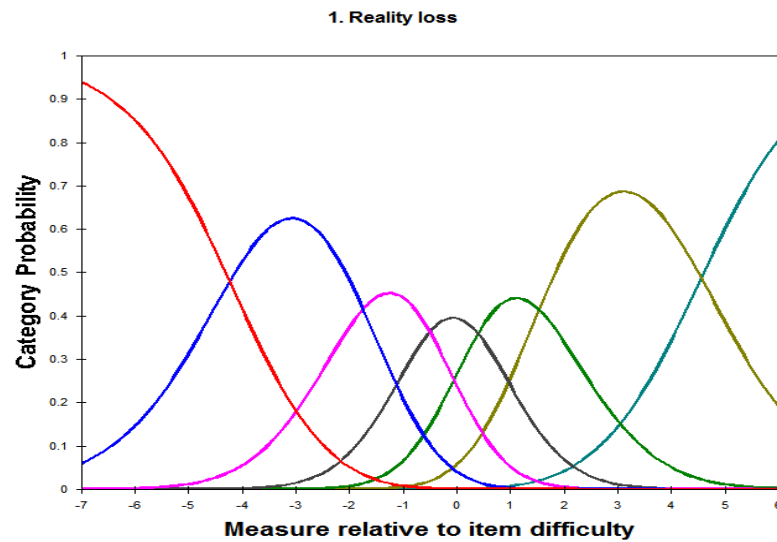


Figure 4.2: The Category Probability Curve of the DELTA scale's item 1: Reality Loss.

4.5.1 Parameters used to report on category functioning

Yan (2009) suggests it makes no sense to start off with the calibration of the items when the categories may be functioning poorly. He suggests poor category functioning must first be remedied before bad items are considered to be omitted. Therefore, as a departing point, the response category structure was first tested, using the basic Partial Credit Model parameters for polytomous scales. They were:

- *Category Probability Curves:* These curves indicate how the response structure is predicted to work for any future sample, provided it worked satisfactorily for this sample. In the Figure 4.2, peaks of the seven categories are all in ascending order along the latent variable of the item. At some defined point on the latent variable; each category in turn is the most probable than any other one of the categories. Furthermore, the cross-over points between the categories are ordered; e.g. the descending curve of each category clearly crosses the ascending curve of the neighbouring category. These cross-over markers are the equal

probability points or the thresholds or the parameters of the CPC. If disordering of the categories occurs, remedial action would be indicated.

Table 4.3: Parameters used in the calibration of category functioning.

| Category Label | OBSVD COUNT | OBSVD AVRGE | OUTFIT MNSQ | Structure calibration |
|-----------------|-------------|-------------------------------------------------|-------------|------------------------------------------|
| Category number | >10 | Expected to advance monotonically with category | <2.0 | Expected to increase with category value |

- *Observed Count* (OBSV CNT): This Rasch parameter in [Table 4.3](#) provides the frequencies of categories been observed in the data set. The observed count also renders evidence towards the Linacre guidelines 2 and 3.
- *Observed Average* (OBSVD AVRGE): This is not a Rasch parameter but an assessment of the sample used to investigate the scale. It is the average of the measures that are modelled to produce the responses as they are observed in the category. The logic is that if the observed averages advance monotonically with the category and the probability curves look good for now for this sample, then one can anticipate the same qualities in future samples [Linacre \(2010\)](#).
- *Outfit mean square* (OUTFIT MNSQ): This Rasch parameter also in [Table 4.3](#) is the average of the outfit mean squares associated in the response of each category. The Linacre guideline 5 suggests values must be < 2.0 to be acceptable for measurement development.
- *Structure calibration*: This Rasch parameter is the points at which adjacent category probability curves intersect, they are not the measures of categories. This point on the latent variable represents the calibrated measure of the transition between two adjacent categories. E.g. the category 4 measure estimates the threshold calibrations between category 4 and the one below, category 3. These points are also called the Rasch-Andrich model thresholds and it is expected to increase with category value. If not, *disordering* of categories are diagnosed ([Table 4.3](#)).

4.5.2 Parameters used to report on item functioning

With the scale raw scores complying with the parameters of assumption and category functioning, the next step was to investigate item functioning. Here again the most commonly used parameters in the literature were considered. The models used by Yan (2009) to analyse and report on item structure were studied. It included: INFIT and OUTFIT MNSQ, Point Measure Correlation, Person and Item Reliability and Variance (Table 4.4)

Table 4.4: Parameters used in the calibration of item functioning.

| Item Labels | Sample size | Category | Infit MNSQ | Outfit MNSQ | PT MSE CORR | Rasch Reliability Pers/Item | Variance Explain by Measure Emp/ Model |
|-------------|-------------|----------|------------|-------------|-------------|-----------------------------|----------------------------------------|
| | | | 0,5 - 1,7 | 0,5 -1,7 | >0,30 | >0.80/-0.80 | > 60%/60% High & close |

- *Infit & Outfit Mean-square (MNSQ)*: This Rasch parameter statistic reports on how closely the scale corresponds to the Rasch model. With values around 1, the measure is considered accurate, meaning the scale's item difficulty range is appropriate to the ability range of the persons under investigation. As the scales are all clinical measures the recommended mean square range for meaningful measurement is between 0,5- 1.7. If the scales overfits (< 0,5), it is interpreted as being too predictable due to dependency in the data, and if it underfits (> 1,7) it is interpreted as being too unpredictable due to the presence of unexpected outliers.
- *Point Measure Correlation (PT MEA CORR)*: The indices should report a noticeably positive correlation of >0.2. This confirms that the distribution and direction from easy to difficult on the scale's latent variables is aligned with the severity or ability of the patients. Rasch expects the

lowest category on the latent variable to be easier for severely ill patients than the highest category.

- *Rasch reliability person/item*: In reliability analysis the RMM quantifies the probability of the scale reproducing the same relative location of the measurement point in future applications given the same patients to observe. RMM reports on both person and item reliability, e.g. a "high person reliability" (>0.80) means that there is a high probability that persons estimated with high measurements actually do have higher measurements than persons estimated with low measurements. The same consideration applies to "high item reliability".
- *Variance explained by measure*: This criterion gives an account of the basic assumption of unidimensionality. It reports both the empirical and modelled values and must be interpreted as follows: If the data fit the Rasch model perfectly, and the raw variance explained on the empirical values is reported as, say, 79.1%, then that number would have been 73.4%, which is reported as the modelled value. However, quality is not interpreted only by how close the empirical and modelled values are, but also how high the percentages are ([Fischer Jr 2007](#)).

4.6 Summary

In Chapter Four the broader scientific methods and considerations of the study to validate the nursing measures were discussed. In each of the following four chapters the analysis and outcomes of each nursing measures are discussed individually. The details of the methods that were applied for each nursing measure are explained in the detail required as it pertains to the individual measure.

CHAPTER FIVE: THE ALPHA

*"Unstable patients are admitted into an acute hospital and are discharged when stable.
Question is: what is stable?"*

HJ Loubser

5.1 Introduction

Independent functioning of bodily organs and systems signal the end of acute curative nursing and the beginning of sub- or non-acute restorative nursing. In the first instance the objective is to "stabilise" the malfunctioning organs and systems, and in the second instance the objective is to restore independence by maximising their patients' activities of daily living. Whilst acute curative techniques require a patient to be immobilised (Daniels 2004), restorative techniques require active mobilisation towards full independence. The determination of the interface between acute and sub- /non-acute is thus of vital importance to determine the switch between the nursing processes. Numerous biometric parameters are used for various ailments to establish if the patient has achieved the status of being "stable" enough to be transferred to the restorative nursing domain. However, establishing patient stability is a complex and highly skilled clinical decision which is usually left to the physicians and the nursing team. The physician uses the laboratory reports and the clinical nurse reviews the nursing reports to arrive at such a finding.

As the finding of "stability" is linked to the acute diagnosis type, so should co-morbidities and pre-existing co-disabilities of the patient also be considered when the status of "stability" can be declared. However, these co-attributes do not always reflect, or are not well represented in the physician's biometric parameters from the pathological laboratories or the radiological investigations. The nurses rely more on the first signs indicating a return of the activities of daily

living, and if immobility persists, nurses construe it as a need for extended acute nursing. In such situations where the patient's biometrics show good functioning of organs and systems, but present with a prolonged inability to mobilise, the acute care nurse may mistake it as a need for ongoing acute nursing care, resulting in prolonged lengths of stay in acute facilities. This is mostly prevalent in geriatric patients, medically complex cases, patients with multiple trauma and complications and also neurological disorders. In these cases there is a high risk that the acute care nursing process might wrongly identify these patients as requiring ongoing acute curative care, with a resulting recommendation to the physician that the patient is "too weak" to be transferred to sub-acute care.

It has been well documented in the literature that early onset of rehabilitation increases the likelihood of a good patient-outcome (Sirios, Lavoie, Clermont & Dionne 2004). Therefore, the sooner the patient is referred for restorative nursing care, the shorter the in-patient's length of stay, the lower the costs of the episode and the more independent the patient is at discharge (Kunik, Flowers & Kazanjian 2006). However, this efficiency is dependent on speedy discharge from acute care (Maulden, Gassaway, Horn, Randell, Smout, & De Jongh 2005), which depends on how accurately the acute team arrive at a decision on the patient's stability. In South Africa the lack of restorative nursing sciences and processes, together with the reluctance of the acute curative nurse to refer patients to restorative nursing, is the main barrier to this efficiency drive. There is no objective measure available to the acute nurses to establish whether the patient is stable enough, from a nursing perspective, to be transferred to sub-acute or non-acute care. Therefore, this study introduces the ALPHA scale to achieve nursing measurement of patient stability.

5.2 Development of the ALPHA

In 2007 the South African Database for Functional Medicine (SADFM) conceptualised, designed and developed a nursing measure to establish patient

stability in acute care. The aim was to explore nursing intuitive knowledge to establish which patient characteristics constitute the interface between acute and sub-acute/non-acute nursing, and how that can be objectively observed and quantified. In other words: how can the curative nurse working in the acute care context quantify whether a patient is stable enough to be transferred to restorative nursing care? With only acute nursing available as respondents and without any formal restorative nursing sciences established in South Africa, the aim of developing an objective nursing measure on patient readiness for transfer to an “unknown” restorative nursing science environment created significant challenges.

Nevertheless, the researcher persevered to establish a nursing consensus on which human functions to observe as indicators of a stabilised patient. By definition, patient stability is indicated by the degree of independent functioning of human organs and systems. However, as most of these are usually measurable and available with the biometric data from special investigation laboratory reports and interpreted by the physicians, the nurses in the acute care settings were not readily forthcoming with additional observational information. As a way out of the impasse, the concept of burden of nursing care was introduced.

The burden of acute care nursing is the inverse of organs and systems' functionality, meaning that the lower the functioning, the higher the burden of care. Therefore, the nursing burden is low with high functioning organs and systems. This is true for the ICU patient on life support systems where the functioning of organs and systems is very low and the nursing burden is very high. As the patient's functioning of organs and systems improve, the nursing burden decreases (e.g. high care instead of ICU). In the general ward the patient is expected to reach levels of independent functioning of organs and systems with resultant low levels of care. In the sub-acute facility less nursing procedures are required to stabilise organs and systems but more managerial skills are required to manage the multidisciplinary team to ensure maximum independence on the activities of daily living. With this concept in mind, the

acute care nurses were more agreeable to participate as collaborators in the design and development of the ALPHA. Facilitated by the researcher they explored the possibilities of quantifying patient stability expressed as a function of burden of nursing rendered to the patient's organs and systems. Care was taken to only include the nursing burden of supporting organs and systems and not including the care rendered to the activities of daily living (eating, grooming, dressing, mobilisation etc) as these are considered restorative nursing items.

Firstly, the twelve major organs and systems of the human body were used to represent the 12 items of the ALPHA (see figure 5.1). Each item was divided into 7 categories with the first category being the highest burden of nursing applied because the patient's organs and systems are functioning at their lowest at this level (e.g. ICU). Each category represents different level of nursing burden of care. Thus, each category infers inversely to the functioning of the organs and systems.

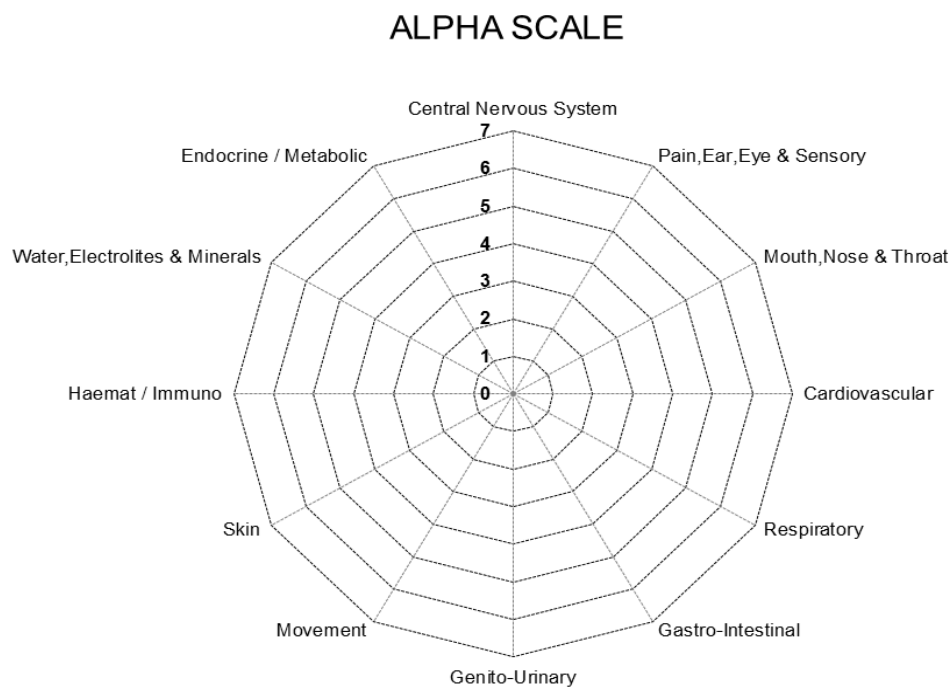


Figure 5.1: Radar graph representing the ALPHA scale structure.

The first option was to use a frequency scale based on the hours of direct nursing rendered per day as an indicator of the burden of care, but this was dismissed as the ALPHA can only score nursing activities rendered to the bodily organs and systems and not to the activities of daily living. Therefore, having to estimate nursing time and separate organs and systems time from time spent on activities of daily living was considered a risk to nursing utility.

Finally, it was agreed that the nursing observations would be based on the nursing services rendered. The respondents developed the basic structure represented in Table 5.1:

Table 5.1: Outline of the ranking of the ALPHA categories.

| Category | Nursing burden of care | Typical nursing environment |
|-----------------|--------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| One | Procedures requiring ICU nursing support | ICU curative nursing |
| Two | Procedures requiring high care nursing support | High care curative nursing |
| Three | Procedures requiring general ward nursing support | General ward curative nursing |
| Four | Procedures required to regain independence such as physical and mental rehabilitation, education, training, dignity etc. | Sub-acute restorative nursing (e.g. rehab, convalescence, end-of-life) |
| Five | Procedures required to maintain optimal functioning | Non-acute restorative nursing (e.g. long-term or home-based care) |
| Six | Procedures required to prevent functional decline | Non-acute preventative (e.g. outpatient care) |
| Seven | No nursing procedures required | Not applicable |

When the framework for the ranking of the categories was agreed upon as a point of departure, it became apparent that numerous other interfaces on the patient severity (or burden of care) continuity might be revealed. Thus, if successful, the ALPHA might provide scores indicating the five interfaces between ICU / high care / general ward / sub-acute / non-acute. It also became evident that the items might be indicators as to where failure is most prevalent. For example: in respiratory system failure the associated affected items might

be the cardiovascular system, but little failure might be found on the endocrine or genital-urinary system. Thus, the respiratory system might at a certain point in time require a very intensive nursing burden of care (e.g. ICU support), whilst the endocrine system might at the same time require no or very little nursing burden of care (e.g. non-acute support). This anticipated anomaly was a matter of concern, but it was unclear how it would eventually be reported on the Rasch model.

A detailed user-manual of the nursing services, procedures and techniques belonging to each category was first drafted for the central nervous system as a pilot study. When the ranking order was understood and agreed upon, the process was duplicated for each of the other eleven items representing the remaining organs and systems. The first draft was submitted to numerous nursing respondents for editing. The final draft was then intended for training, testing and implementing the ALPHA in a private general hospital facility where ALPHA admission, intermediate and discharge scores were to be recorded on all acute adult patients admitted (>18 years) into the ICU, high care and general wards.

5.3 Conceptual framework of the ALPHA

The ALPHA measure is presented as a new nursing measure to monitor patient severity as reflected in the functioning of the bodily organs and systems. It is intended to be an observational scale rendering objective cross-cutting scores on the severity of the patient irrespective of the diagnosis or underlying pathology. Structurally the ALPHA is an intermediate length measure with 12 items, each intended to have a hierarchical rating consisting of seven categories. Thus the ALPHA's total summed raw score range is between 12 and 84 with the lower scores indicating increased severity of the ailment, meaning the lowest levels of organs and systems failure requiring the highest levels of nursing supportive services. No patient self reporting is required for the ALPHA.

Its objective schedule of scoring is anticipated to be familiar to all professional nurses. However, nurses need to undergo training and testing for proficiency in applying the ALPHA before administering the scale. The training takes 2-3 hours.

5.4 Qualitative study: nursing utility

To prevent any bias, a new acute hospital facility was selected, where no nurses participated in the design or development of the ALPHA, for the sampling and data collection. The ALPHA was thus a new concept and experience for the nursing staff. The hospital and nursing manager agreed to test the ALPHA and 15 professional nurses representing the ICU, high care and general wards were selected to attend a 2-3 hour training session. The nurses were then requested to implement the ALPHA and to collect and record patient severity data daily on a hard copy for four months. Regular meetings were held to ensure compliance with data collection, however, within the second week the nursing staff defaulted in scoring patients and collecting the data. Neither the researcher, nor the nursing manager or the hospital manager could motivate the nurses to implement the ALPHA.

The main reasons why nurses defaulted in implementing the ALPHA and collecting data were summarised as follows:

- Implementing the ALPHA was considered an additional nursing administrative task with few benefits or usefulness to the nursing process.
- More than 70% of patients recover spontaneously (e.g. cold surgery cases) and should not be subjected to ALPHA scoring and recording as their recovery to full independence is predictable.
- It is not the nurses' decision to triage patients to other levels of care; it is the physician's prerogative. Therefore, the ALPHA evidence is irrelevant to the nurse.

- The concept of restorative nursing was not clear, and until the validated evidence-based outcomes of the restorative nursing model are available, acute care nurses will continue to implement the curative model of nursing until the patient has reached the appropriate levels of independence in the acute facility. In other words, the acute nurse did not see the need for restorative nursing.
- Quantifying the transition between the various nursing levels was determined by the nursing utilisation ratios and the parameters set by the funding industry. This was considered to be adequate at the time.
- In the acute hospital the patient-evidence based parameters are the concern of the treating physician, not of the acute care nurse. The acute nurse is task- orientated and not patient-outcome orientated.

With the above reservations expressed by the registered nurses, the ALPHA patently failed the nursing utility tests. As nursing utility is the first and foremost logical consideration when starting the validation process of a routine nursing measure, the complete rejection of the nurses to implement the ALPHA was considered a dead-end in the validation study of the ALPHA. However, the researcher did persevere and data of the ALPHA was collected over a period of six weeks as described in the quantitative section of the study below. However, informal interviews on the nursing utility of the ALPHA with the single nursing rater did concur with the objections raised by her colleagues as described above.

5.5 Quantitative study: construct validity

5.5.1 Sampling and data collection

In agreement with the nursing and hospital manager, a single PRN was recruited to be trained and tested in the application of the ALPHA to collect daily patient data in ICU, high care and in a mixed general ward. Funding from the

SADFM was made available to remunerate the rater nurse for the extra hours to collect the ALPHA scores over a period of six weeks. The nurse was supervised by the researcher. The scores were done daily and recorded on hard copy. The researcher collected the scores regularly and entered it into an Excel spreadsheet.

For the quantitative analysis, ALPHA data were collected on 759 adult patients admitted into an acute hospital over a period of six weeks. All this data were collected daily by the single PRN, who scored all adult patients admitted into the facility, irrespective of diagnosis or underlying pathology. No exclusion criteria based on gender, race or ethnicity prevailed. All admissions, daily, intermediate and discharge ALPHA observations were recorded, totalling 2367 raw ALPHA scores.

5.5.2 Data preparation for the Rasch analysis

The ALPHA is designed as a polytomous observational scale to score the functioning of human organs and systems. In its raw format it is an ordinal scale meaning that its data cannot be summed into total scores and therefore not be used in secondary analysis. To render meaningful statistical information, the ALPHA has to be transformed from an ordinal to a linear interval scale, and the RMM was proposed to perform this transformation.

However, not all types of ordinal data are suitable for analysis with the RMM. To make valid inferences from the analysis, the data have to meet certain requirements of the RMM, such as complying with four basic assumptions, namely uni-dimensionality, local dependence, monotonicity and non-intersection category probability curves ([Iramaneerat, et al 2008](#)). The preliminary RMM analysis and assessments of the raw ALPHA data showed that it did not satisfy these basic requirements to be calibrated by the RMM. Although in-depth studies on the RMM might be feasible at a later stage,

it did not fall into the ambit of this study; more so with the failure of the ALPHA to show any nursing utility properties. The Alpha scale thus failed both validation criteria.

5.6 Conclusion

The validation of the ALPHA is a perfect example of where nursing intuition concurs with mathematical precision. The nursing utility of the ALPHA was outright rejected by the nurses as not being useful, and although the researcher persevered by collecting 2363 ALPHA responses on 759 patients with dedicated accuracy, the Rasch model outcome concurred with the nurse's outcome. The ALPHA is not only an additional burden to the nursing process; it also is nothing more than an ordinary rating scale. Its data is worthless to the entire healthcare industry. The most probable reason for this failure of the ALPHA in this particular study is that the latent trait explored with the ALPHA; the functioning of organs and systems, fall within the scientific measurement realm. It is not measurable within the human sciences which require observational scores on the relationship between the ability of a person versus task difficulty. The Rasch model confirmed this distinction very clearly in this study.

CHAPTER SIX: THE BETA

"I think if you do something to somebody that can do it for him or herself, I think you are not giving better care"

Care-giver in focus group (CG: p 13. 9-11).

6.1 Introduction

The consequences of trauma and illness may be impairment, disability or handicap (WHO 1980) and the key to understanding the extent of these phenomena is to accurately document the activities of daily living (ADLs) (Lundgren Nilsson 2006). Measuring ADLs provide evidence concerning how people live with such functional losses. The Uniform Data System for Medical Rehabilitation (UDSmr) is the originators of the Functional Independence Measure (FIM) with the motto: "As we function, so shall we live" (Granger 2011). It emphasises the need for accurate measurements to appropriately address the nursing needs of patients living with temporary or permanent functional loss. Measuring ADLs is a complex task, and many rating scales have been introduced since the 1980s. However, in the USA the FIM has become the gold standard of ADL measurement (Nilsson, Sunnerhagen & Grimby 2005) when the Centres for Medicare & Medicaid Services (CMS) included it in 2002 into their mandatory Inpatient Rehabilitation Facilities-Patient Assessment Instrument (IRF-PAI 2001). To qualify for the CMS Prospective Payment System (PPS) rehabilitation facilities had to register and implement the IRF-PAI. Thus, by implication, in the absence of FIM data, rehabilitation facilities in the USA will not receive funding from CMS.

By taking the lead in measuring ADLs globally, the FIM received extensive

positive and negative literature reviews. Its construct and predictive validity was scrutinised and the literature widely reported that the outcome of the FIM was dependent on the training received, the sampling done and the analytic methods used (Lundgren Nilsson 2006). However, very little was reported on the FIM's nursing utility and little is known on how well, if at all, the nursing profession have embraced the FIM as a routine nursing measure. The FIM was designed in the 1980's by a multidisciplinary team and promoted to be used by clinicians, and this approach was never changed. Although nurses in the USA do specialise to become nursing clinicians (AANP 2011), this speciality accreditation is not available in South Africa. In any case, clinicians, whether physicians, nurses or therapists are not the primary caregivers offering routine ADL support to patients. Clinicians may periodically score patient ADLs in therapeutic environments, resulting in scores that might be regarded as having a bias towards patient potential abilities. However, if clinicians require the patient's actual performance reflected on how the patient performs routinely in the presence of the caregiver, they will infer scores by interviewing the nursing assistant on the patient's abilities. Thus, the researcher concluded that the FIM has an inherent structural defect; it is designed for clinicians but requires valid information that can only be provided by the primary care givers.

Over and above quantifying patient outcomes, nurses also find ADL measurements particularly useful in calculating the level of nursing care required (McGillis-Hall, Doran, Baker, Pink, Sidani & O'Brien-Pallas 2001). As the inverse of functional ability indicates functional inability which equates to the burden of nursing care, the measurement of ADLs is a promising rationale to calculate nurse staffing in restorative nursing facilities (Heinemann, Kirk, Hastie, Semik, Hamilton & Linacre 1997). If ADL measurements can be done routinely (daily) by nurses as an integral part of the nursing process, the data would be available to calculate the burden of care. Therefore, if routine functional data were available, levels of staffing could be calculated either in advance or in real-time for each individual facility unit, rather than the traditional retrospective staffing models using a cross-sectional aggregated norm across similar units

that may have diverse patient mixes often requiring case-mix adjustments. Furthermore, availability of routine functional data would also assist in moving away from the traditional use of negative staffing indicators (e.g. deaths, co-morbidities complications) towards a more positive focus on staffing appropriately for measurable patient outcomes. (Lang, Hodge, Olsen, Romano & Kravits 2004).

With nursing staffing as the main consideration, Nelson, Faan, Powell-Cope, Palacios, Luther, Black *et al* (2007) published an article providing significant insights into the restorative nurse's needs to quantify ADLs. However, it is the collateral information provided by Nelson *et al.* (2007) on the FIM nursing utility that is of interest to this study. As the FIM rating scale is the mandated ADL measure by the IRF-PPS since 2002, Nelson and co-workers assumed the nurses would provide the daily FIM data for the researchers to develop the prospective staffing model based on functional outcomes. This did not turn out to be the case (Nelson *et al* 2007).

The Nelson study was designed, in collaboration with the developers of the FIM system (UDSmr), to ask nurses to collect FIM data routinely every day on all their patients for 30 consecutive days. The purpose of the study was to calculate the link between patient functionality and the required burden of nursing care. At the time, 806 rehabilitation facilities in the USA were accredited to use the FIM system, meaning they all received training and have been tested to the point where 80% of the clinical staff achieved 80% and higher marks in the FIM credentialing examination. The Nelson study however, required a computer generated randomised sample of 806 facilities, set by region. Finally, the sample provided 235 rehabilitation facilities to participate in the nursing study. However, the nursing management of 75% of these sampled facilities declined the invitation to participate as it was thought to place too high a burden on their nursing staff. Only 54 facilities agreed to participate in the daily FIM scores for 30 days. Of these 54 facilities the mean registered length of

nursing rehabilitation experience was 8.11 years (7-15) and only 17% were certified rehabilitation nurses. Their nursing managers gave them an above average proficiency rating of 3.81 (1-5) in rehabilitation nursing. In spite of these nursing qualifications and long standing experience in the use of the FIM, it was concluded by Nelson's research team, which included the original accrediting agency UDSmr, that the rehabilitation nurses would require retraining and re-accreditation in order for them to provide accurate FIM data. In spite of all the endeavours above, an average of 9,48% of the FIM data were missing daily at all facilities.

This is evidence that the nurses, however well intended and committed to comply with the FIM program to provide routine data, could not maintain the daily effort to observe the functional changes of the patient, to convert them into a score and to record the score. At this point, it is important to keep in mind that functional gain is the primary focus of rehabilitation nursing and that the collection of FIM data on the patients' actual performance is not only mandatory by the USA funding system, but that the inverse of the FIM scores is also theoretically measuring the burden of nursing care. Routine FIM data should thus provide core nursing process information to provide patient-based-evidence of applying efficient nursing staffing ratios. The FIM data should thus render vital information for the restorative nursing processes; however, the cumulative evidence in Nelson's study indicated that the nursing process and the FIM were not compatible. Either the nurses fail the FIM or the FIM fail the nursing utility.

The researcher introduced the FIM into South Africa in the late 1990's and trained, tested and accredited more than one thousand nurses and rehabilitation therapists working as multidisciplinary teams in 84 sub-acute facilities over a three year period. The aim was to introduce the FIM as a routine nursing tool, but the results were disappointing. The registered nurses found the concept of routinely scoring patient function interesting; but not having made the

transition from the curative to the restorative nursing model, they found having to score ADLs out of the scope of their practice. In keeping with the curative nursing model, they preferred to rely on their intuition and record ADLs broadly in the narrative in patient files. However, when scrutinising their descriptive reporting, it showed little awareness of the need to use the basic ADLs as indicators to monitor restorative nursing processes. Their references to ADLs were minimal, and when recorded, the information was subjective, incoherent, incomplete and inaccurate. The standardised format of their nursing documentation was still formatted for a curative nursing outcome, which consists of long checklists to ensure that the nursing tasks were completed satisfactorily. The need to collect patient-evidence based FIM data was consequently seen by sub-acute nurses as an additional burden to their checklist driven curative nursing process. The nurses thus rejected both the concept of collecting FIM scores as well as using the FIM as a nursing measure.

With these prevailing barriers the researcher approached the nurses in focus groups to explore the core reasons for rejecting to record ADLs in their nursing process. Their five basic concerns preventing them from implementing the FIM were as follows:

- Rendering assistance with ADLs is not a professional registered nurse's primary function; it is the primary function of the nursing assistants. The registered nurse supervises and manages the process.
- The registered nurse is not in a position, neither is it in her scope of practice to observe, score and record ADLs routinely; the nursing assistants are in the best position as they are supporting the patients with these tasks.
- The ADLs are considered the domain of the therapist, not the domain of the nursing profession. In the multidisciplinary team the therapist must take responsibility for FIM scoring.

- The registered nurses cannot teach the FIM to the nursing assistants as the FIM contains language and concepts that professional nurses are not fully familiar with. Training, testing and accrediting of nursing assistants to score and record accurate scores is thus out of the question.
- With the above four concerns made clear, the final conclusion was that only under duress would the nurses implement the FIM and under these circumstances the nurses would refuse to accept accountability on the accuracy of the scores. The FIM was considered a cumbersome and inappropriate additional burden on the nursing profession.

Thus, the final verdict on implementing the FIM was: South African nurses found the FIM having no nursing utility and therefore would not voluntarily participate in the multidisciplinary team to routinely produce FIM scores. The cumulative result on this was significant:

- Routine ADL scores are not possible as therapists do not have the continuity of patient contact to produce such scores.
- Actual ADL scores are not possible as therapists only have a limited window of observation while a patient is in a therapeutic environment. This is when patient potential is measured.
- With no immediate prospect of formal training in restorative nursing for the South African nurse, the opportunity of skills transfer from therapeutic team members to nursing team members also seems to be lost if nurses cannot participate in the ADL scoring process.
- Sub- and non-acute patients requiring functional improvement are not likely to benefit from the South African nursing practices which are unable to measure their outcomes. These patient outcomes will remain unpredictable.
- Health care processes will not have accurate routine data available to set up databases to manage the clinical governance of the sub- and non-acute facilities of care.

- The limitations outlined above were pointed out to the nurses in meetings with the researcher. An attempt was made to convince them of the importance of nurses collecting ADL data. Finally it was agreed that an ADL nursing score could be tested if it had a nursing utility which satisfies the following objectives:
- It should be accepted that observing, scoring and recording ADLs are done at nursing assistant's level and not by professional nurses. Thus, the ADL nursing measure must be trainable, testable and recordable at that level of nursing.
- Only if it is successfully implemented at nursing assistants' level, will the registered nurse be able to supervise the process.
- Supervision of the nursing assistants in collecting the scores daily will be the task of the nursing process, but the supervision of the accuracy of scoring must be shared with the therapists.
- Nurses should participate in the design and development of the ADL nursing score, provided that it is tested by the nursing assistants.
- To prevent confusion and disorder within the multidisciplinary team, the proposed ADL nursing assistants' score must have the same format as the FIM; but for trademark considerations must be called the BETA nursing measure.

6.2 Development of the BETA

Based on the above principles the design and development of the BETA began in 2005 and the first BETA version was produced with the participation of registered nurses. Their input was used to facilitate four main transitions from therapeutically orientated FIM to the nursing orientated BETA. Firstly, attention was given to simplifying and focussing the definitions of the Items without changing the construct of measurement. Secondly, the algorithms of arriving at a category were changed to fit the mind-set of the nursing assistant without changing the score value. Thirdly, the rules were amended to match the nursing

assistants' scope of practice and level of participation. Finally, and most probably the most challenging, was the BETA's conversion to become a routine nursing score and not a periodic therapeutic score.

The first version was tested in 2006 on NAs in a training session with their PRN supervisors in attendance as observers. The NAs were informed that it was a first version and they were invited to verify the appropriateness of the terminology contained in the BETA draft, as well as questioning its practical implications on their scope of practice. The first result was very positive in terms of motivation to participate. Numerous changes were made and valuable suggestions received, indicating that the PRNs originally participating in the design and development of the BETA were not fully aware of the detail and nuances in the NAs' primary care-giving scope of practice. The evidence of the NAs' depth of knowledge and insight took the registered nurses by surprise, indicating the richness of the intuitive awareness regarding human function vested in the NAs.

The second version was tested later in 2006 and again insights from NAs brought valuable adjustments. This process continued until version ten was completed in 2008 and the PRN's were in agreement that the NAs could start collecting data. Since then, NAs have been collecting ADL data routinely under supervision of PRN.

There was however an important caveat in the BETA development. Nurses reported that three items in the original FIM are problematic for a nurse to observe directly and score appropriately, namely transfer into a bath, walking / wheelchair locomotion and climbing stairs. In South Africa all well designed nursing facilities are without baths. It has been replaced with wheelchair-showers for the patients that cannot yet walk and walk-in showers complete

with chairs inside for those who can walk. Thus, transfer into or out of a bath does not occur, and transferring into or out of a shower only occurs when the patient is mobile enough to walk into the shower. It is thus not possible for nurses to observe patient ability across the latent trait range of the item difficulty: "Transfer into and out of bath/shower". A similar case comes about with stair climbing. Modern facilities are designed without steps to prevent patients having to climb stairs. This activity is therefore impossible to observe routinely. Furthermore, if stairs are available somewhere else in the facility, it is considered a therapist task, not a nursing function, to observe, test and train patients on stair climbing. Thirdly, the walking / wheelchair item has both an algorithm to rate ability and a frequency to detect distance achieved. These two parameters had to be combined into a single score and nursing assistants found this difficult to score without reverting back to the manual.

The researcher was thus notified that nurses might not be able to observe these three activities routinely as originally intended, and thus would not be able to make actual observational scores available. To arrive at a routine daily score they would have to simulate situations or guess or predict the scores. The challenge to the researcher was first to establish through Rasch how accurate the nurses are scoring the "bath/shower transfers", the "walk / wheelchair" and the "stairs" and based on this information, make a decision to include or exclude these items in the BETA nursing measure.

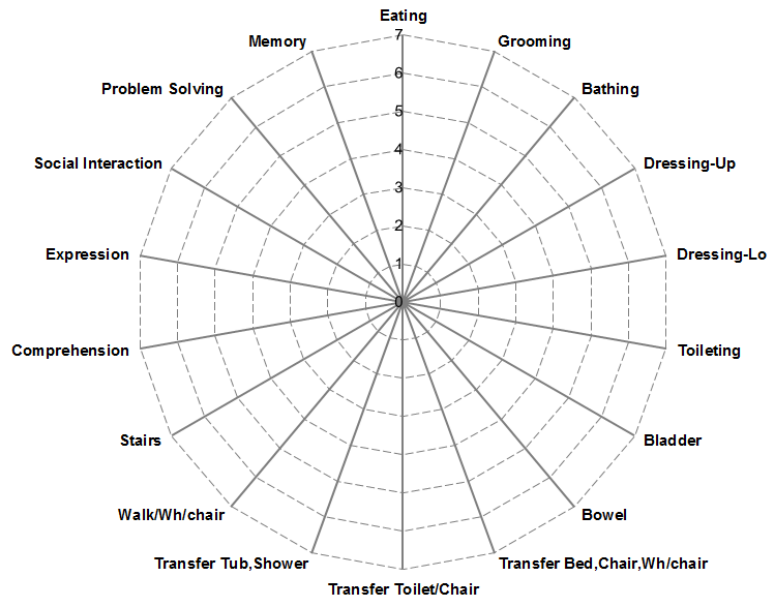


Figure 6.1: Radar graph representing the BETA scale structure, based on the FIM platform.

6.3 Qualitative study: nursing utility

6.3.1 Sampling and data collection

For this section of the study, data were required from nurses who have used the BETA to observe, score and record their patients daily for more than six months. However, to avoid bias, it was decided not to collect data from one of the facilities that have participated in the development of the BETA. Therefore, a new facility was identified to test the BETA's nursing utility. The BETA was also re-tested in a non-acute environment where predominantly caregivers and not NAs implemented the BETA scoring. The assumption was that if caregivers can successfully use the BETA to observe, score and record patient functioning daily, the higher qualified NAs should easily achieve the same performance. Thus, a 48 bed geriatric frail care non-acute facility, with 45 permanent caregivers working under the supervision of 6 PRNs, was selected to re-test the BETA's nursing utility. The whole compliment of the permanent nursing staff

were trained, tested and credentialed to use version 10 of the BETA scale. Their task was to daily observe and score patient functioning.

Six months after implementation in the geriatric facility, two focus group interviews were held to explore the nursing utility experienced. The two focus groups were homogenous; the first only with PRNs, and the second only with caregivers. The PRNs were requested to respond to questions pertaining to whether the BETA is adding value to the nursing process, and the caregivers were requested to report on the ease of the BETA application and the benefits of BETA use within their scope of practice.

On the particular day the focus groups was scheduled, the professional registered nurses and caregivers on duty for that particular day were invited to participate. Thus, a non-probability purposive sampling technique was used. The staff availability on the scheduled day was three professional registered nurses for the first focus group and five caregivers for the second focus group. The groups were held at the facility in an allocated room familiar to the respondents. Respondents were put at ease, the reason for the interview was explained, as well as, the ethical procedures to be followed, and their rights for withdrawal from participation. Upon agreement of the above, both interviews were conducted, recorded and transcribed.

6.3.2 Data analysis

The same design and techniques of deductive content analysis described in Chapter Four applied to the BETA study. As the goal of a deductive content analysis was to conceptually validate or extend the existing assumption of nursing utility, it provided the researcher with a focused approach, limited to the four questions originally posed during the development phases. These four questions also provided a framework to predetermine the coding scheme for analyses. In other words, the original four questions posed to the PRN

collaborators during the design and development were anchored as the categories of this analysis. (Hsieh & Shannon 2005).

6.3.3 Results from the professional nurses' focus group

The references in brackets (e.g. PN: p 6: 12-18) refer to the page and line numbers in the transcriptions where the evidence was found. The acronyms PN and CG specify the transcription of the registered professional nurse and caregiver focus groups respectively.

It was generally agreed amongst the professional nurses that whilst the traditional nursing care plan included care tasks regarding the skin, bowel, sleep, feeding, comfort, and keeping free of pain, the BETA has opened a different nursing window of observing a patient. The original nursing care plan focused on what the nurse has to do for the patient; whilst the BETA focused on what the patient can do for themselves. As nurses scored and quantified patient ability, they have found a new medium of communication amongst themselves to better understand patient ability and to find solutions amongst themselves to improve it. This has changed their goals in quality nursing care. Before, it related to the amount of tasks done for the patient now, quality nursing relates to how much independence they can restore for their patients. Patient independence has become the topic of their discussions, which the nurses experienced “as a positive development” (PN: p 6: 12-18).

There was significant evidence that the BETA is acting like a nursing care plan which guides the caregiver into new insights what to do with the patient. As they score patient functioning, caregivers know exactly what their patients' abilities are. The recorded scores serve as continuation documentation between shifts and are discussed at handovers. Previously, the caregivers were very task orientated, doing all the activities for the patients, but when scoring with the BETA scale, caregivers have become aware that the patients are capable of

doing many of the tasks themselves. The BETA algorithms also guide the caregivers to the next scoring level, and as a result the caregivers are challenging the patients to achieve a higher scoring level. The result is the caregiver and the patient is motivating each other towards higher levels of independence resulting in increased BETA scores. Furthermore, increasing the patient's BETA scores has now become a commodity representing a personal achievement for the caregivers. The BETA thus motivates the caregivers to increase patient independence. If the scores did not improve, the caregivers consulted the professional nurses regarding techniques to improve the scores. For the professional nurse the routine *"BETA scoring is a very big help"* (PN: p 2. 12).

The nursing organisation at the facility allowed the carers to score their patients daily on a hard copy where after the weekly score sheets are submitted for electronic data capturing. The professional nurse daily discusses the changes in the patient score sheets with the care givers. For instance, if the patient has not had a bowel action for the day the "bowel management" score for the day is left open. The *"omissions and changes in scores form a very good discussion"* around the patient between the professional nurse and the caregiver (PN: p 5. 20-21).

The professional nurses established that patient ADLs are within the scope of practice of the caregivers and that the BETA scoring method is becoming second nature to them. The caregivers can observe functional ability of the patient and recognise change as it occurs. In fact, they are more aware of how patients function than the professional nurses are. With the caregivers' new founded perspective on restoring patient independence, the overall quality of nursing has definitely improved (PN: p 6. 24). There is very strong quantifiable evidence that the patients' (frail aged persons) BETA scores have improved substantially since the carers have been implementing the BETA, indicating improvement in patient independence. These changes relate to quality-of-life improvements which indicate patient-evidence based quality of nursing (PN: p 3. 4-8, 16-19).

In addition to the caregivers' new founded skill as BETA scorers, there was also a noticeable amount of assertiveness, accountability and increased motivation. This new role has taken the caregivers out of the housekeeping team as they are now recognised and integrated into the nursing team. They have become enthusiastic about their new role and it has become a morale boost to them. They now participate in the nursing discussions and find the new language in expressing themselves by means of scores when discussing patient functioning easy to manage. They understand their task as that of implementing a nursing procedure and they do it with the required discipline for which they are praised appropriately. Most importantly, their achievements are now quantifiable (PN: p 4. 4-7, 15-23).

The caregivers are now more aware of the comprehensiveness and importance of maintaining patient abilities than previously. Their attitudes toward their patients and their job have also changed as they have become more involved in restoring patient abilities. In fact, they feel that they are actually the leading part of the team when it comes to improving patient functioning. *"They are enthusiastic, wants to lead the team and doing very well. They are definitely more alert and aware of what the patients can do and cannot do. They really know their patients from A-Z. ... it is wonderful!"* (PN: p 6. 14-23)

An interesting secondary observation was that professional nurses have found that they can use the BETA to make an assessment of caregiver performance. There was consensus that if a caregiver has problems in scoring a patient on the BETA, it is because they do not have the intuitive awareness to assess a patient, or do not show enough interest to observe their patient's ability, or cannot function as a caregiver. A review of the BETA scores immediately shows who is capable of doing the work and who is not. *"We can score our care-givers' (according to their competency as a carer) according to their scoring"* (PN: p 8. 7-20)

The final comment: The BETA can be embedded as a routine nursing measure into the nursing process. *“It is working very well with us. If it can work here it can work anywhere else”* (PN: p 8. 2-4). The care plan and the scoring are integrated into one process; e.g. what you do every day, you score every day. Recording the correct score makes the difference. It is a very positive experience, however, it requires control from the nursing manager and the registered nurses to support and guide the caregivers during the first few months. The daily scoring and recording makes a great difference. It helps them to continuously verify the scores amongst each other as they go about their daily caregiving with the patients. During the first few months all scores were achieved informally through consensus discussions amongst themselves. The PRN's reported that *“it was very encouraging”* to witness the caregivers enthusiasm to learn more about scoring and improving patient functioning. (PN: p 9. 16-18).

6.3.4 Results from the caregiver's focus group

Caregivers claim that the introduction of the BETA scale changed their perception of their scope of practice. Previously, they believed it was their task to do everything for the patient, even if the patients were capable of doing basic tasks for themselves. Feeding patients that could eat themselves were considered amongst the routine tasks of the caregiver (CG: p 3. 1-4). *“Spoiling patients”* by doing everything for them was considered excellence in caregiving (CG: p 3. 6-7). *“Before, what was on our minds was that it was our job to do everything for our patients, even if they can do it for themselves, we must do it for them. In our minds that was our job.”* (CG: p 15. 1-5). *“This is how we were taught as caregivers to look after our patients ... to do everything for them”* (CG: p 13. 15- 18). As a result, patients become increasingly spoilt, demanding and abusive as their experiences of hopelessness mounted. This made the task of the caregiver very difficult and tiring as there is no change for the better ... it only gets worse every day. As a result, they reported that their patients were getting quieter and even stopped talking, socialising, eating, walking, etc. (CG:

p 3. 11-17). In this environment the caregivers reported a feeling of not being in control of the situation, becoming either emotionally involved or detached, and also being physically tired and having aching backs from lifting and transferring people (CG: p 16. 20-25) ... *"The job was hard before"* (CG: p 16. 14).

However, with the introduction of the BETA scores on patient ability, the caregivers reported a new mind-set on the concept of caregiving. Initially they experienced the training and implementation as contradictory to their job-description which is being task orientated. To make this change they needed continuous re-confirmation from their nursing superiors. It was also difficult to establish a BETA score while working and to apply one's mind at the same time to find a method, strategy or technique of how to improve the patient's independence. The changes to move away from the usual daily task tick list (feeds, bed, baths, transfers, grooming, dressings etc) towards a patient outcomes score sheet of how much the patients can do for themselves, was also a daunting challenge. For continual confirmation and support, caregivers required strong nursing supervision and constant reassurance. However, they reported a smooth transition within two weeks from a task driven process to a patient outcomes mind-set (CG: p 10. 25).

"I think if you do something to somebody that can do it for him or herself, I think you are not giving better care" (CG: p 13. 9-11). This piece of evidence sums up the new awareness that became prevalent amongst the caregivers within a month of using the BETA. They repeatedly mentioned this new mind-set as the conceptual framework of their new approach to caring for their patients. They are also confident that the BETA has substantially improved the quality of their care to their patients. They can now quantify the quality of their care in their patients' improved BETA scores. Moreover, they also anecdotally confirmed improvements in patients that they never thought possible such as: *"we see many (improved) changes in our patients"* (CG: p 3.12-17); *"(they are) getting better, better really"* (CG: p 14. 12); *"yes, our patients are getting more independent now"* (CG: p 14. 17-19); *"also their memory is improving and they*

are expressing better” (CG: p 17. 10-12); *“there is a lot of improvement”* (CG: p 17. 18).

Previously, the caregivers thought their patients’ on-going decline was irreversible, but now they have discovered a restorative remedy that they have control over; and this empowers them. They felt they have gained a lot of experience in a short period and they are enjoying their job more than before (CG: p 8. 9-10 & p15. 16-18). With patients now making some physical effort with some of the caregivers’ strenuous tasks such as the transfers, they experienced less back injuries (CG: p 15. 20-25 & p 16. 18-22).

However, the transition of the focus of care also created barriers. Some patients became set in their ways of being served and refused, although very capable, to be retrained to become independent. However, patients behave differently with different caregivers’, e.g. some male patients refuse to do self-care tasks when there are female caregivers that can do it for them. None the less, the caregivers are discussing these behavioural deviances and have found agreeable solutions amongst themselves to overcome these attitudes. (CG: p 8. 7-10).

There is consensus that *“it does help to score every day because you can see the changes in the patient from day to day”* (CG: p 3. 22-23). Routine scoring is particularly helpful if the caregiver requires monitoring to test the outcome of a new restorative technique they have implemented. It is also helpful to detect subtle improvements in independence due to quality caregiving. But in the field of gerontology sudden declines may indicate the onset of a clinical emergency e.g. dehydration or flu. With routine scoring the caregiver awareness is maximised and they notice daily changes easily and score accordingly (CG: p 4. 19), which benefits the patient and adds value to the scope of practise of the professional nurses.

Finally, there is strong evidence that the BETA scores are successful in creating a universal language amongst the caregivers. In fact, the caregivers mentioned that during the first months when they discussed patient scores during *"lunch time, tea time and even went to bed with scores"* (CG: p 10. 20-25), They also created sessions as a group amongst themselves to discuss scores and work sessions to score difficult patients. Furthermore, they also initiated discussions on restorative techniques to increase their patient scores. The scores were also discussed with the nursing staff and any changes were reported to the professional nurse on duty.

6.4 Quantitative study: construct validity

6.4.1 Sampling and data collection

For this section of the BETA study, data was collected over a period of four years from those sub-acute facilities which participated in the development of the BETA scale. Only data collected from version six to ten were used in the analysis as these changes were more cosmetic than structural. All these facilities were fully accredited in the application of the BETA. The whole nursing staff were provided with a training manual, trained and tested. Accreditation certificates were issued when 80% or higher was achieved by the nurses, and data was collected from a facility when 80% of the nursing staff was accredited.

All admissions into the facility were observed, scored and recorded within 48 hours. Intermediate score changes were recorded as they occurred and were presented as a nursing progress report at weekly multi-disciplinary meetings. All discharge scores were recorded on the day of discharge. Each nursing team designed and developed their own nursing process documentation to record the BETA raw scores, and an electronic web-based application was provided to import the admission, intermediate and discharge scores from the nursing documentation.

All adult patients (>18 years) admitted into the facilities were included in the study, irrespective of their diagnosis or underlying pathology. No exclusion criteria based on gender, race or ethnicity prevailed. All admission, weekly intermediate and discharge BETA observational scores were pooled, totalling 16,639 raw BETA scores representing 5356 patients.

6.4.2 Data preparation

The aim of the BETA analysis was to confirm if the Beta data in its present scale format satisfy the basic goodness-of-fit Rasch requirements, allowing informed recommendations on the future of the BETA as a nursing measure. The first concern in the data was local dependency, as the total raw scores contained admission, intermediate and discharge responses for most patients. This was overcome by using a 15% computerised random sample done in Excel with the selection based on the frequency distribution of the total admission, intermediate and discharge observations. Therefore the final dataset for analysis had 4235 raw score observations representative of admission interim and discharge scores free of local dependence.

This raw score dataset, free of local dependency, was used for the first BETA analysis across all 18 items (see figure 6.1). It produced unsatisfactory results as disordering of categories were observed across the first 13 motor items. The 5 cognitive items however provided better results than the 13 motor items. The motor and cognitive items were then grouped into a motor and cognitive sub scale, and the analyses repeated. The 5 cognitive items showed marked improvement when analysed separately as a sub-scale; the 13 motor items also improved, but disordering still remains a problem in this sub-scale. This spontaneous improvement in the category ordering of the cognitive sub-scale function, led the researcher to consider whether the 13 item motor scale can benefit from further subscale analysis. In re-considering whether the 13 motor

items should be calibrated as a coherent scale or divided into sub-scales for meaningful routine nursing observation; the following came to mind:

- the concern from the nurses that they were not able to observe some items (e.g. stairs, bath transfers)
- the difference in rating scale structure between the items (e.g. counting frequencies, using Likert measures, using algorithms or using a combination such as walking/wheelchair where distance and ability should be brought into consideration which was causing debates amongst nursing raters),
- the inability of nurses to observe some responses on certain items routinely (e.g. counting frequencies in bowel and bladder control),
- the same observations for different items (e.g. dressing lower body and toileting) causing structural local dependency problems.

Considering the above nursing concerns, clinical knowledge and the Rasch reporting on category disordering, a decision was made to create a further sub-scale structure for the 13 motor items (see Figure 6.2). The four Beta sub-scales are referred to as the self-care-, toileting-, mobility- and cognitive subscales. From here onwards the four subscales were each calibrated separately. As Verhalst and Glass (1995) stated, there are two methods that scale developers could use to enhance measurement construction namely: to omit “bad” items and/or temporarily remove the observations that clearly misfit the Rasch model. Linacre (2010) suggested clinical observations with under fitting responses over 1.7 MNSQ logits are usually associated with careless mistakes. This data is too unpredictable for Rasch measurement development, and could be removed for calibration purposes. Therefore, the most miss-fitting data (< 1.7 MNSQ logits) in the study were removed, leaving each sub-scale with its own data set free of under fitting data and free of local dependence.

6.4.3 Rasch calibration

Following the [Verhalst and Glass's \(1995\)](#) directive, the calibration was started by omitting the malfunctioning items. Based on both nursing knowledge and the Rasch information, the following item adjustments were made to the four sub-scales (as set out in [Figure 6.2](#)):

- The toileting item was relocated to the sphincter sub-scale. This was done as some of the observed activities in dressing lower body and toileting are the same (e.g. pulling pants down and up, and loosening and fastening zips, buttons or belts whilst steadying) and in this instance the Rasch model requirements of local dependency might be violated. By moving the toileting item to the sphincter control sub-scale, a nursing scale for measuring the complete toileting experience now becomes feasible.
- The stairs item were removed from the mobility sub-scale as it is neither working as a Rasch rating scale (e.g. disordered categories) nor does it satisfy as a routine nursing measure (e.g. nurses refuse to take patients up and down stairs).

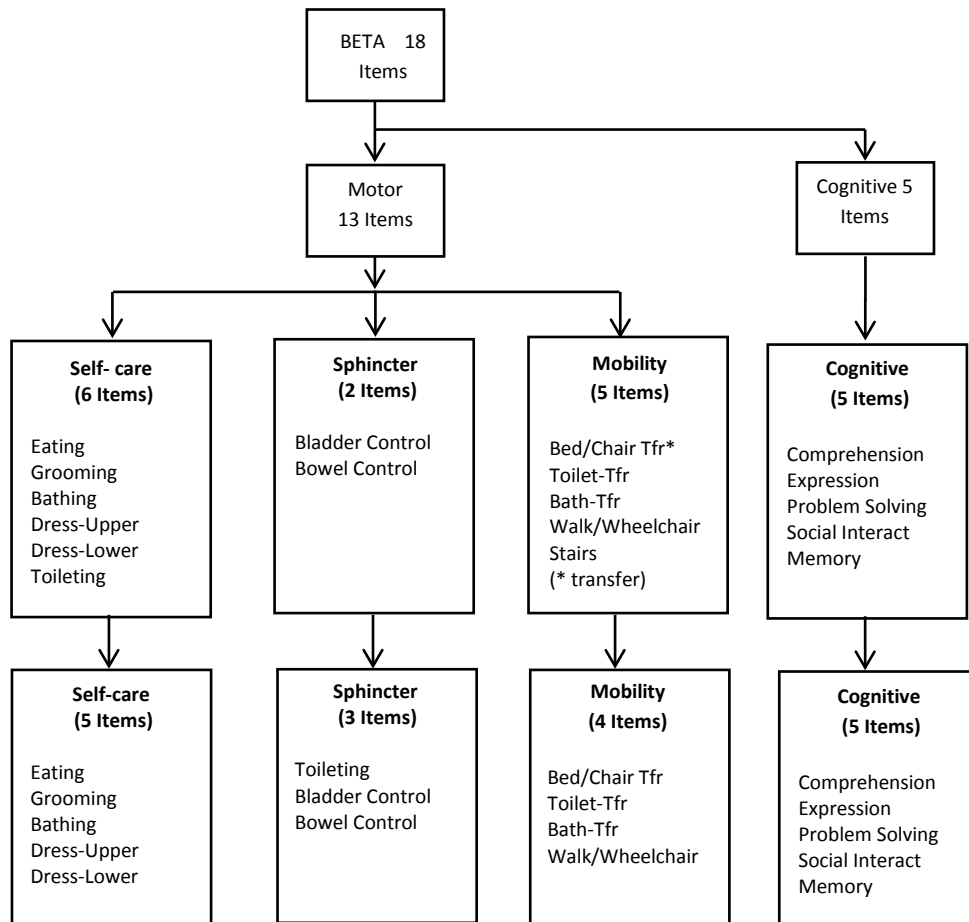


Figure 6.2: Development of the four BETA sub-scales.

To test the dependability of the researcher's construct theory, the variable or item person map (see Figure 6.3) on the BETA Self-care subscale was analysed. In Chapter Four the interpretation was explained. From Figure 6.3 it is evident that in the BETA Self-care subscale the person ability matches the item difficulties. However, significant ceiling or floor effects are revealed. This can be clinically explained as follows: A substantial amount of the BETA data analysed were admission scores when the patient was transferred from an acute nursing settings into rehabilitation nursing setting. On admission the patients has no, to very limited, functional ability, and this is most evident in patients with brain injuries, high spinal injuries, those with complex medical conditions, etc. Some of these patients never regain any of their self-care abilities. The ceiling effect is explained with certain patients recovering quicker from the items on the self-care subscale than the other BETA sub-scale items. Therefore patients may be fully functional in the self-care items, but still receive restorative nursing for the remaining BETA sub-scales.

scale, the category probability curves in some items were submerged by others causing disordering which obstructed meaningful calibration. This deficiency indicated that category collapsing was needed in order to obtain an interpretable category structure for some items. The guidelines of [Linacre \(2004\)](#) were followed in the process of combining adjacent categories. This, amongst others, was to check that the outfit mean squares should not exceed two logits, and threshold advances be at least 1.4 logits for a three-category scale or one logit for a five-category scale. In the final draft, the collapsed category structures also satisfied the category probability curves need for having ordered intersections with neighbouring curves. The new item rating scale structure for those items requiring remedial collapsing of their categories, is summarised in [Table 6.1](#). The remaining items' categories were functioning well.

The “New structure” column in [Table 6.1](#) must be interpreted as follows: The original structure for all items consisted of seven categories in the 1234567 order. If Rasch arrived at a conclusion that nurses could not distinguish satisfactorily between two neighbouring categories (say 2 and 3) and suggested that these two categories would function better as one category, then they were collapsed into one category. The new structure of the item would now read 1223456 meaning that categories 2 and 3 were collapsed to form category 2 and thereby reducing the item's total category structure into 6 categories.

[Table 6.1](#): The new structure of the BETA categories that required collapsing.

| Subscale | Item | New structure | New categories |
|-----------------|--------------------|----------------------|-----------------------|
| Self-Care | Grooming | 1233456 | 7 reduced to 6 |
| Toileting | Bladder Control | 1223334 | 7 reduced to 4 |
| | Bowel Control | 1223334 | 7 reduced to 4 |
| Mobility | Bed/Chair transfer | 1233456 | 7 reduced to 6 |
| | Toileting transfer | 1233456 | 7 reduced to 6 |
| | Bath transfer | 1223345 | 7 reduced to 5 |
| | Walking/Wheelchair | 1223345 | 7 reduced to 6 |

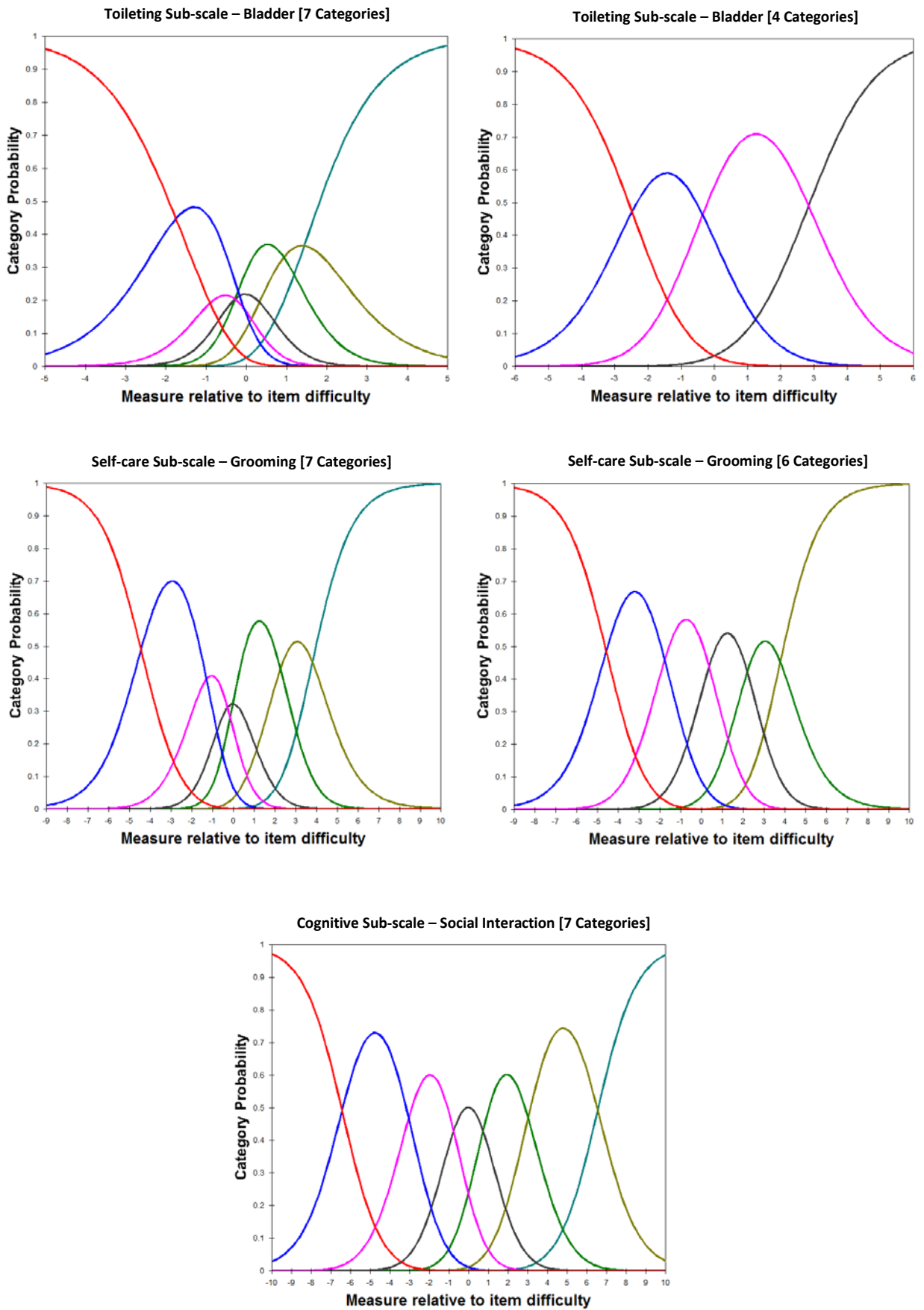


Figure 6.4: Examples of the BETA category probability curves before and after collapsing.

After collapsing of items, only the mobility sub-scale's walking/wheelchair item reported an outfit MNSQ value of 3.41 logits which, according to the Linacre (2002) guidelines, was too high for meaningful measurement. However, during this calibration it was decided not to delete this item from the mobility sub-scale, but rather to recommend re-visiting the category definitions of the walking/wheelchair item. The nurses have reported difficulty in arriving at a score when taking both distance and ability into consideration. This has clearly been identified by the Rasch model and should be addressed at a later stage, but not in this study. All the other sub-scales showed reasonable to very good compliance with the Linacre guidelines for quality measurement properties.

6.4.5 Results on item functioning

With the category functioning satisfying the Linacre (2004) guidelines, verification on the Rasch fit statistics parameters for item functioning was required. The Rasch model selected for reporting on the fit statistics for each sub-scale were the Infit and Outfit MNSQ values, the Point Measure Correlation (PT MSE CORR), Rasch reliability for person and item, and the variance experienced by measure (Table 6.2). According to the literature these parameters are the most widely referred to and commonly used.

The Infit and Outfit MNSQ values are the core statistics reporting on whether the scale fits with the Rasch model or not. It also indicates how closely the scale appropriates the Rasch model. When values are around one logit, the measure is considered accurate. However, for clinical scales such as the BETA sub-scales, Linacre (2010) suggested Infit and Outfit MNSQ value ranges between 0.5 and 1.7 as reasonable for quality measurement. Thus, with both Infit and Outfit MNSQ values ranging well between these indices on all four sub-scales, it can be concluded that the item difficulty range is appropriate to the ability range of the persons under investigation Linacre (2010). Consequently, all four of the BETA subscales can be regarded as measures with good levels of accuracy

and predictability. In the case of the cognitive sub-scale the values range well between Fischer's (2007) quality criteria for excellence, being $< 0,77 - 1,3$.

Table 6.2: Results on the BETA item functioning.

| Sub-Scales | Item Labels | Categories | Outfit MNSQ | Outfit MNSQ | PT MSE CORR | Rasch RELIABILITY Person/Item | Variance explained by measure Emp / Mod |
|--------------|--------------------|------------|-------------|-------------|-------------|-------------------------------|-----------------------------------------|
| Self-care | Eating | 7 | 1.54 | 1.49 | 0.89 | 0.97 / -0.95 | 86.3% / 85.9% |
| | Grooming | 6 | 0.94 | 0.93 | 0.92 | | |
| | Bathing | 7 | 0.85 | 0.83 | 0.90 | | |
| | Dress-up | 7 | 0.68 | 0.66 | 0.91 | | |
| | Dress-lower | 7 | 0.94 | 0.87 | 0.87 | | |
| Toileting | Toileting | 7 | 1.45 | 1.34 | 0.94 | 0.99 / -0.93 | 76.2% / 76.6% |
| | Bladder | 4 | 0.67 | 0.66 | 0.96 | | |
| Mobilisation | Bowel | 4 | 0.83 | 0.81 | 0.96 | 0.99 / -0.99 | 87.8% / 86.6% |
| | Bed/Chair Transfer | 6 | 0.68 | 0.62 | 0.97 | | |
| | Toileting Transfer | 6 | 0.57 | 0.51 | 0.97 | | |
| | Bath Transfer | 5 | 1.28 | 1.19 | 0.95 | | |
| Cognitive | Walk/Wheelchair | 5 | 1.38 | 1.49 | 0.95 | 0.99 / -1.00 | 88.1% / 87.9% |
| | Comprehension | 7 | 0.90 | 0.85 | 0.97 | | |
| | Expression | 7 | 0.93 | 0.86 | 0.97 | | |
| | Social interaction | 7 | 0.86 | 0.88 | 0.97 | | |
| | Problem Solving | 7 | 1.04 | 1.04 | 0.97 | | |
| | Memory | 7 | 1.12 | 1.11 | 0.97 | | |

The PT MSE CORR reported a noticeably positive correlation of > 0.3 . This Rasch evaluation confirms that the distribution and direction from easy to difficult on each of the BETA sub-scales' latent variables is in alignment with the severity of the patients. Rasch expects the lowest category on the latent variable to be easier for severely disabled patients than the highest category.

The Rasch reliability for person and items quantifies the probability of a BETA sub-scale to reproduce the same relative location of the measurement point in future applications, given the same patients to observe. RMM reports on both person and item reliability, e.g. a "high person reliability" means that there is a high probability that persons estimated with high measurements actually do have higher measurements than persons estimated with low measurements.

The same consideration applies to “high item reliability”. All four the BETA sub-scales obtained significant person reliability and item reliability values. Three sub-scales are well into Fischer’s (2007) range of “excellent” quality in item and person reliability (> 0.94). The toileting sub-scale falls into Fischer’s “very good” classification with an item reliability value of 0.93.

The variance explained by measure is the Rasch criterion for dimensionality and reports both empirical and modelled values. It must be interpreted as follows: if the data fit the Rasch model perfectly, and the raw variance explained on the empirical values are reported as 86.3%, then that number would have been 85.9%, which is reported as the modelled value. However, quality is not only interpreted by how close the empirical and modelled values are, but also by how high the percentages are. According to Fischer (2007) values higher than 80% and as close together as the reported values in the Self-care, Mobilisation and Cognitive sub-scales, indicate “excellence” in quality in measurement properties. The toileting subscale values of 76.2% / 76.6% again falls within Fischer’s category of “very good” quality.

6.5 Conclusions

Strong implicit and explicit evidence supported the BETA’s nursing utility to facilitate a restorative nursing process. It has made the nurses aware of the value of having the activities of daily living embedded into the restorative nursing process. This was manifested explicitly in the ease with which the nurses reported the primary caregivers to recognise, observe and express the BETA scores in their daily routine; and implicitly with the apparent new awareness that improvement in BETA scores has a direct correlation with improvement in the patient’s independence and resultant wellbeing. Thus, as they became aware that restoration of patient independence is their core nursing focus, they actively explored techniques of increasing the patients BETA scores. In this process the primary caregivers also became aware that they are in control of nursing effectiveness.

The Rasch model clearly illustrated that the BETA is not functioning as a singular scale with 18 items, but rather functions as a suite of four well-defined sub-scales working in harmony to measure and explain four different dimensions of the activities of daily living. In retrospect, these sub-scales also make clinical pragmatic nursing sense as the sub-scales are linked to the nursing activities performed by the same nurse at the same time. For example, the toileting sub-scale will record the nursing activity of undressing lower body, followed by the bladder and/or bowel sphincter control and the cleaning at the end as one activity observed by the same nurse at the same time. It will not be experienced as three different activities to be observed by different nurses. There is a perceived value to the nursing profession by using four individual measuring units rather than one all-inclusive tool. With more specific measures, the nursing assistants might be better equipped to realise and monitor new restorative nursing techniques in overcoming specific disabilities such as in transfers. Although the four sub-scales should be analysed separately, their summed totals might still reflect a total BETA on the patient's ADLs.

This study is limited to the basic Rasch analysis to verify if the BETA has potential as a valid nursing measure to collect patient outcomes data routinely. This objective has been achieved with success.

CHAPTER SEVEN: THE GAMMA

7.1 Introduction

All longitudinal studies exploring the outcomes of restorative nursing should include the instrumental activities of daily living (IADLs). The ability to live independently in one's chosen environment, usually one's own home, is an aspect of functional assessment that has been overshadowed by scores reflecting the activities of daily living (ADLs). This, in the therapeutic sciences of rehabilitation, includes the well-established scores of ADL as found in the Barthel ADL Index, the Katz ADL Index and the Functional Independence Measure (FIM). Most of the research in the area of IADLs continues to be in the sciences of elderly population, looking at the ability of individuals to live in their own homes.

The first IADL scoring framework was published in 1969 by Lawton and Brody in their classic article: *Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living* using the phrase "instrumental activities of daily living" (IADL). They did not provide a definition for IADL, but instead described a scheme of competence into which behaviours would fit, taking self maintenance (ADL) as the lowest level. Behaviours that indicated successively more complex levels of function were ascribed to the IADL scale. The Lawton and Brody IADL scale includes necessary activities but is not limited to domestic chores, household management, outdoor activities, and transportation. Restorative therapy or community resources are directed to these activities to allow the disabled or the elderly to remain in their homes.

Following Lawton and Brody's seminal work, different studies composed their

own set of different instruments measuring one or two aspects of frailty under investigation. Also different institutions have developed their IADL scales to serve their institutional needs, an example being the [RIC-FAS \(1987\)](#). Others used the concept of assessment and adjusted it to their needs. Some needed to examine specific patient populations and in the process changed the evaluation methods of tests. In the process even the test items and the scaling of categories differed from test to test. As these specific IADL scales were developed for specific impairments and specific objectives and used by specific professionals they developed their own phrases for their adjusted scales as “extended ADL” ([Nouri & Lincon 1987](#)), “social ADL” ([Wade, Legh-Smith & Hewer 1985](#)), “advanced ADL” ([Reuben, Laliberte, Hiris & Mor 1990](#)) and the list may not be exhausted.

Nonetheless, the Lawton and Brodie’s phrase “IADL” has withstood the test of time and is generally well accepted as the descriptor of independent living items and categories. However, although there is a conceptual understanding of IADL; there is still no agreement as to the exact categories or items to be included in an IADL to serve as a generic clinical outcome measure supporting a large-scale national database. In the most recent literature the quest for standardised IADLs information to be incorporated when analysing functional data is becoming more prevalent.

IADLs are human functions that underpin and sustain more nuanced and complex social activities than the ADLs. The original Lawton and Brody IADL items included the abilities to use a phone, do shopping, prepare food, do housekeeping, do laundry, manage own transportation, medication and finances. The IADLs are considered the minimal daily functions a person must be capable of to live independently. Losing one or more of these activities increases the risk of living independently. This correlation between IADL measures and the ability to live independently unlocks the possibility for the nurse practitioner to measure and quantify the nebulous concept of frailty from

its onset and through its progression. Although the term frailty is mostly associated with the end stages of the ageing process, it also refers to the disabled and to other vulnerable people who is not necessary old (Brink 2010).

According to Pel-Littel, Schuurmans, Emmelot-Vonk & Verhaar (2009) nearly all studies published in the field of clinical frailty report deterioration in IADLs. They report the most common characteristic associated with elder person's frailty is an unstable physical and mental condition and diminished physical reserves. When these reserve capacities are exhausted the decline in daily functioning manifests itself. Different authors refer to this critical point where the balance is disturbed and refer to it as the point where the frailty threshold has been exceeded (Markle-Reid & Browne 2003). Thus the extent to which a deterioration of IADLS is detected becomes a determinant of the onset of frailty in elderly persons.

According to Lynch (2004), physical aging is associated with a progressive loss of muscle anabolic hormones and growth factors leading to loss of muscle mass (sarcopenia), a slowing of movement, and a decline in strength. These factors increase the risk of injury from sudden falls and leads to growing dependence by the frail elderly on assistance in accomplishing even the most basic tasks required for independent living. The debate exists as to whether these intrinsic changes are immutable or reversible. There is clearly a profound need for restorative strategies that can slow the effects of ageing on muscle function, and restore muscle size and strength in the frail elderly so that their quality of life can be maintained or improved.

The consequences of frailty are usually measured in terms of mortality, morbidity and institutionalisation. The frail population, due to the ageing process or disability, are vulnerable to increasing dependency and a consequent higher risk of being admitted into nursing care facilities (Bandeem-Roche, Xue, Ferrucci, Watson, Guralnik, Chalves *et al.* 2006). They risk a lower quality of life than their non-frail peers (Strawbridge, Shema, Balfour, Higby & Kaplan 1998)

Their unstable physical and mental condition and diminished physical reserves also leave them vulnerable to illnesses and at risk of injuries leading to acute nursing care settings (Van, van Geel, Geusens, Kessels, Nuwenhuijzen-Kreusenman & Brink 2008). Pel-Littel, Schuurmans, Emmelot-Vonk & Verhaar (2009) reported an absence of a gold standard to measure frailty, meaning it is not yet possible to comment on the validity and reliability of the diagnosis of frailty. Although a limited number of total frailty tests have been developed little is known about their validity.

The consequences to the nursing profession of not being able to measure the physical frail characteristics mostly relates to the planning and managing of nursing services to the frail population. In South Africa the gerontological nursing services have a structured approach in supporting the independent living needs of the elderly persons. This usually starts with periodic home visits to the independent living older persons to assess their needs for supportive services. These nurses are known as the “cottage nurses” in the retirement village industry. If nursing support is required and rendered the older person to continue living independently, the person is considered receiving “assisted living nursing services”. But when the burden of rendering assisted living services reached a distinctive peak, the older person is triaged - depending on the need - to either frail or Alzheimer institutionalised living nursing services. Finally, the end-of-life care needs are taken care off with the emphasis of dignity preservation. These levels, settings and services are all managed by the nursing profession and based on rendering functional support to the elderly person. However, without a validated measurement of elderly peoples’ functional abilities to live independently, the triaging of an elderly person between these interfaces remain a subjective issue and often become a matter for debate between the elderly, their families, the nurse and the management of the retirement village. This is more so when the higher costs of rendering assistive living and frail care are taken into consideration.

A validated objective nursing measure that would provide evidence that an

elderly person needs to move from an independent living unit to an assisted living unit or from an assisted living unit to an institutional frail care would ease the emotions and tension accompanying the triage process. It would make an ill-defined subjective process quantifiable. With routine nursing measures the situation over time become statistically predictable and decisions can be pre-empted well in advance. Interim nursing strategies, techniques and interventions will delay the inevitable time for making triage decisions, and the success of these nursing processes will be quantifiable.

In the matter of younger persons with chronic disability or residual permanent disability after rehabilitation the same nursing process prevails and the nursing profession is confronted with similar barriers. A validated measure embedded routinely in the nursing process would achieve similar benefits.

7.2 Development of the GAMMA

The Gamma was conceptualised, designed, developed and implemented in 2007 to address the concerns of the community-based “cottage nurses” working with persons at risk of becoming frail. The researcher invited experienced nurses to participate in the process of creating the GAMMA. The community nurses’ implicit brief to the GAMMA developers was thus to provide an objective observational measure that would enable them to arrive at an accurate score to quantify their client’s ability to continue living independently. The measure must enable routine longitudinal scores to track change in person functioning overtime. Thus the difficulty level of the scoring system must be simple, allowing also the caregivers working in assisted living to score and record daily the functioning of the client as they observe it.

The development was a qualitative approach whereby the researcher presented the community-based nurses individually and in groups with the classic Lawton Instrumental Activities of Daily Living Scale ([Lawton & Brodie 1969](#)) as a frame

of reference. They were asked to comment on the appropriateness of introducing the Lawton IADL Scale in their work place. Their initial objection was the Lawton IADLS do not conform to their needs as it was a self-report scale requiring 15 minutes to complete by their clients, who might be offended if given such a self-rating scale. This presented a challenge to the researcher to convert the Lawton self-reporting questionnaire into an observational rating scale by selecting some of the items in the Lawton IADLS together with some other items considered appropriate.

The eight items selected were meal preparation, household chores, home/car accessibility, commuting, errands, money matters, self-medication, and emotional stability. These items were tested for all possible situations across the diverse South African populations, living environments and social statuses. The Lawton IADL item “ability to use telephone” was purposefully omitted as it would implicate a section of the aged or disabled population negatively for not having the financial resources to own a telephone or cell phone.

With these eight items in place the community-based nurses were required to identify the possible levels for each item starting with the lowest level associated with a total inability to perform the activity. When consensus was reached on the description of level one, the next task was to identify level two. This required a description of level two plus a nursing observational or probing method to distinguish between level one and two. This process continued until level seven was established as the level of full functioning. The nursing observational or probing methods were analysed by the researcher to arrive at an algorithm whereby the training manual would guide a novice to arrive at an objective score.

Two perceived barriers to the GAMMA were identified. The first is the situation where the person does not perform an activity, having never done so in their life but is fully independent because someone else is doing that activity for them. An example is a husband who enjoys the daily meals his wife is preparing for

him, but has never cooked a meal in his life. The second barrier is the person that lives in an environment that does not necessarily demand them to perform that specific GAMMA activity such as managing money matters; how does one score such a person? The questions therefore posed were: are we scoring actual performance or potential performance? A solution was found by requiring the community nurse to score and record both actual and potential scores for each client. However for this study only the actual scores were analysed.

7.3 Conceptual framework of the GAMMA

The GAMMA is introduced as a new community based nursing scale to be used with clients living within a high risk of becoming frail. Its aim is to render objective cross-cutting scores on the ability of the client to continue living independently. Structurally the GAMMA is a relatively short scale with eight items, each having a hierarchical algorithm arriving at seven categories. Thus the total summed score range is between 8 and 56 with the lower scores indicating the inability to live independently. No client self-reporting is required. Furthermore, the GAMMA claimed to be an objective framework of observation and rating, a familiar process and experience for community nurses to get acquainted with. Training and testing for proficiency in the GAMMA lasts 4-6 hours.

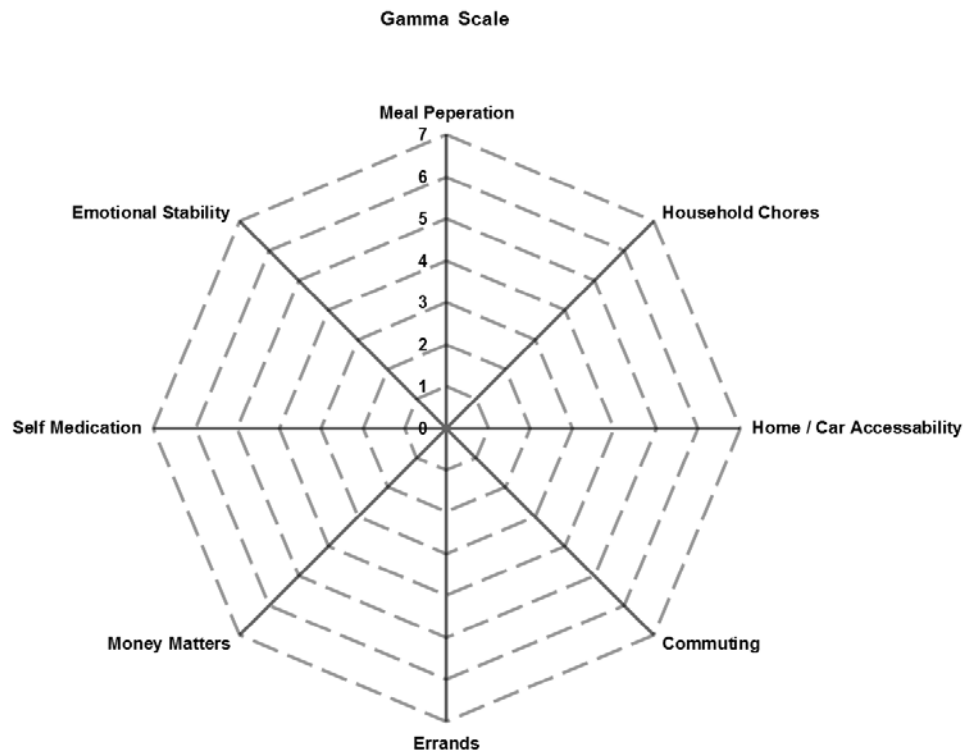


Figure 7.1: Radar graph representing the GAMMA scale structure.

7.4 Qualitative study: nursing utility

7.4.1 Sampling and data collection

For this section of the GAMMA study, descriptive data was required from nurses who were not participants in the development of the GAMMA, but have used the GAMMA to observe, score and record their patients. Nurses working in retirement villages and having a working knowledge of their residents' independent living abilities were asked to participate. Firstly, they were issued with the GAMMA training manual, then trained until they fully understand the GAMMA logic, and thereafter requested to implement the GAMMA in their retirement village. The objective was to arrive at a cross-sectional baseline score for all their residents living in independent and assisted living units, but

soon longitudinal scores were recorded as they observe changes in GAMMA functioning.

Originally the sampling design only included the cottage nurses whose primary task it is to visit the residents routinely for medical check-ups and to review their medication. However with the introduction of the GAMMA, the nursing managers took a keen interest, to the extent that they did the scoring. They also believed themselves to have a better holistic view of the residents' abilities to perform the GAMMA items. This was an interesting finding, as it would have been thought that the cottage nurses would be better placed to observe patient functioning. There were exceptions where cottage nurses did the scores, but even then, it was within collaboration with the nursing managers. They were provided with a web-based software application whereby they could enter the scores and graphs of the GAMMA profiles could be generated.

Five months later a focus group session was held with four nursing managers and one cottage nurse to provide evidence on the usefulness of the GAMMA to the nursing care-plan and process in retirement villages. They were all professional registered nurses. The focus group was held in one of the facilities' boardroom. The nurses did not know one another but were introduced and given adequate time to familiarise with each other. The broad questions related to whether the GAMMA is useful to the nursing care-plan, does it improve the quality of nursing, can it be embedded into the nursing process as a routine measure, and does the scores create a universal language amongst nurses regarding patient functional status?

7.4.2 Data analysis

Similar to the BETA, the GAMMA was also developed with the collaboration of experienced nurses and hypothetically the GAMMA also has a high degree of nursing utility. For this reason a similar deductive approach to content analysis

was followed using the same four broad questions to arrive at a coded framework for analysis. As a result the original four questions posed to the PRN developers to affirm nursing utility were now anchored as the four categories for deductive content analysis (Hsieh & Shannon, 2005).

7.4.3 Results of the professional nurses' focus group

- Usefulness

The nurses agreed that the GAMMA scores and graphs *“really give you a picture of the residents' ability”* to continue living independently (p.5:22-23). Importantly, is that the GAMMA *“helps to evaluate the person with a holistic approach as it shows how the person functions as a whole being”* (p3:2-3). As a result it also gives the nurses a new *“awareness of the needs of the resident”* to enable them to continue living independently (p 2:23). The ability to empirically express this status and to empirically monitor the changes over time in a resident facility is of particular significance (p3:17).

With this quantifiable information now available, the nurses have become aware of their own shortcomings in rendering the appropriate supportive services to address patient needs. Nurses have become aware that they will have to *“totally move away from the old way of nursing and caring”* (p2:5-6) and start looking at what abilities the person have and find a new techniques and care-plans of *“trying to help the person to continue in a process of being able to help himself for as long as possible”* (p2:7-9). Nurses have also discovered that by continuing with the traditional methods of nursing care they were *“giving help without specifically doing something to slow the process of decline”* (p2:11-12). It is now clear that they should move away from the ever increasing nursing support to rather focus on early intervention with restorative methods and techniques to *“keep them independent for as long as possible”*(p2:15). Evidently nurses have become aware that in geriatric nursing their main focus can no longer only be *“structured around the medical side, maintaining hygiene*

and manage medication” (p2:16-17). This will have to include techniques to secure independent living for as long as possible.

Following on the above evidence, the respondents concurred that the GAMMA is useful and contributes significantly to the nursing care-plan and process. Particularly useful is its ability to early identify preventative problem areas that can be resolved with pro-active restorative nursing allowing residents to maintain their independence longer. The GAMMA is a valuable tool for the “cottage nurse” as it *“makes her look at (aware of) the various segments of what is going on in the independent and assisted living scenarios” (p10: 13-14)* Previously the items addressed by the GAMMA were overlooked, taken for granted, or nurses were not even aware of its importance to promote independent living.

The caring professional nurse’s motto is: *“I want to help you” (p10:22)* but then they wrongly identify nursing tasks to make the person dependent on the nurse. The GAMMA has changed this mind-set. With the comprehensive view of the GAMMA, the nurse is now forced into a paradigm shift of considering what nursing techniques the patient requires in maintaining their independence. For this reason *“the GAMMA should form an integral part of the nursing process” (p10:23-24).*

The respondents however warned that by nature nurses are *“very task orientated, and the difficulty is going to be to turn the staff away from this” (p3:20-21),* towards a restorative approach in maintaining independence. However, with proper training, support and supervision, they envisage that the implementation of the GAMMA will benefit everyone. Currently, every nursing day is inundated with nursing tasks, and as the residents get older these tasks increase. The value of the GAMMA vests in its ability to diagnose the primary causes of oncoming frailty early and it also points out the area where early restorative support can be rendered to maintain independence for as long as possible. Training of nursing assistants and caregivers should not only be

focused on how to use the scale, but more importantly, on *“looking at what the scale is pointing out to you and trying to better that (weaknesses) rather than listing it as a new nursing task”* (p4:8-9). To entrench this *“new way of thinking”* (p4:11) within the existing colleagues might be a challenge but, if successful, it would reduce the nursing burden of perpetual task rendering. The respondents’ consensus was that the GAMMA will facilitate this process and *“at the end it is going to be easier on all the staff from not being task oriented”* (p4:5).

- Quality of nursing

The respondents were determined that, if implemented correctly, the GAMMA will facilitate major improvements in the quality of nursing of older persons. If nurses merely observe and record the GAMMA indicators and not react upon it appropriately, nothing will come of improving nursing quality. It requires a multi-task team including the non-nursing staff members such as the observant gardener, cleaner, reception staff, kitchen lady to the administration person and most importantly the manager. It also has to include the resident and their family members. *“The simplicity of the GAMMA includes everyone”* (p12:4). Everyone must be aware, observant and report to the nurses when they observe any changes. However, it is the nurses’ task to score, record and respond appropriately and timeously.

- Uniform language

Interesting and unforeseen evidence was provided on the betterment of the communication and relationships between the nursing professionals and the village managers and trustees. Nurses have found the GAMMA scores to be a useful language in communicating meaningfully with the managers and village trustees about the aging residents’ abilities and needs. The nurses reported a sudden clarity and interest shown by the managers and trustees. This new simplistic intelligibility have motivated managers to request that all new prospective residents are screened on the GAMMA and the profiles are used to establish and communicate the level of care to be provided. The nurses also

reported significant buy-in amongst managers to use the GAMMA as the point of departure at routine meetings with the nursing staff. Furthermore, the managers also request GAMMA scores and profiles as an information briefing before entering into family meetings on resident issues. (p12:16-22)(p13:1-5) (p13:7-17) The GAMMA scores also seems to help the manager to have a better understanding of the nursing complexities, and thereby initiate a more reassuring relationship when nurses have to implement new and creative restorative techniques to facilitate independence (P14:.23-24)

In some retirement villages caregivers had already been given an informal GAMMA introduction. High levels of acceptance and job satisfaction amongst the caregivers were reported. With their new intelligence, caregivers continuously communicate and test their scores with the nursing staff (p12:3-4), report when they observed increase scores as a method of seeking acknowledgement (p12:4), and seemed eager and ready to implement the GAMMA officially. *“They are like a sponge ready to draw up the knowledge”* (p15:8)

- Suggestions

In the last question for suggestions on improving the GAMMA, a concern was raised regarding the safety of residents. They noticed an absence in the GAMMA structure to score and record the basic activities of daily living that would inform nurses of inherent risk factors such as walking, bathing and transferring. As these basic activities of daily living are fully dealt with in the BETA scale, and as these particular respondents are not yet familiarised with the BETA, the concern was dealt with after the focus group meeting. However, of note on this issue was the expressed need that the “cottage nurses” would require, at least in some instances, the BETA to serve as an adjuvant to the GAMMA in completing the full picture of independent living.

The GAMMA item that scores self-medication creates a problem. The step-wise

logic of person ability to function independently includes a score whereby nurses would dispense their weekly pillboxes with medication. The respondents made it clear that the Act 101 of 1965 on dispensing of medication by nurses do not allow nurses to fill weekly pillboxes. Nurses are therefore forced to dispense in daily pillboxes, an unnecessary task which needlessly reduces the GAMMA's independent functioning score of the resident. As it stands now, the GAMMA score 4 on self-medication cannot be utilised as Act 101 is forcing nurses to fill the daily boxes, thus arriving at score 3. This irregularity in logic might impact on the validity of the GAMMA. The nurses also warned that they are finding it difficult to observe decreasing abilities in handling of own financial affairs.

At the end there was a suggestion to consider collapsing the two items pertaining to transport, e.g. home care accessibility and commuting, into a single item. The nurses also remarked that the emotional stability item is not an easy item to observe and might warrant further definitive consideration.

Finally, a nursing professional in the position as village manager provided the following conclusion: *It seems that the GAMMA "brings the quality (of nursing) care to its rightful position. Quality care can now be processed, managed and maintained in a very scientific way. It is important as it forms an integral part of the nursing process. You get quality care that you can measure and that is very important. The results you get with the GAMMA axes allow you to maintain and enhance quality care ... and you can calculate it"* (provide empirically proof).
(p2:18-24)

7.5 Quantitative study: construct validity

7.5.1 Sampling and data collection

For this section, observational data on two homogenous groups of persons

were pooled into a single dataset for analysis. The first group consisted of 428 older persons' functionality in seven retirement villages. Only those living independently in their cottages and those living in an assisted living environment were included in this count. The residents in frail care units were excluded. The senior nursing staff of four of these retirement villages participated in the development of the GAMMA. However, none of these residents "cottage nurses" who collected the data participated in the GAMMA development. The "cottage nurses" or "village nurses" are dedicated to the task of visiting the residents routinely in their homes to observe and render support where needed. The GAMMA was a new experience to all of them and they were accredited in the application of the GAMMA. This included training and testing with the help of a training manual. They then set out to observe, score and record all independent and assisted living residents in their villages. Originally it was intended to be cross sectional scores done on residents as a baseline for future longitudinal studies; however some nurses did follow-up assessment as they observed change in patient functioning. Thus both cross sectional and longitudinal observational scores were done on some residents rendering a total of 468 responses in the retirement village grouping.

The second group of data was collected on 334 patients receiving home based care by a home based care agency nurse. The home based care agency nurse was trained tested and accredited to use the GAMMA, and she scored patients longitudinally on admission, intermediate and at discharge. In total she collected 689 responses. Patients were referred to the agency by medical schemes for convalescent care after an acute hospital or rehabilitation episode of care. None of these patients were residents of the seven retirement villages in the first group. All adult patients admitted into the home based care program over a period of one year were scored. No exclusions were made based on any criteria except age (<18).

The data of both groups were collected on hard copy and in most cases entered by the nursing services into a web-based software application. The rest were faxed to the researcher for capturing. The pooled raw data from both groups totalled 1157 responses.

7.5.2 Data preparation

The Rasch analysis was aimed to establish if the GAMMA satisfies the basic goodness-of-fit Rasch requirements, allowing informed recommendations on the future of the GAMMA as a nursing measure. The first concern in the data preparation was local dependency as the total raw scores contained follow-up responses in the retirement village grouping and admission, intermediate and discharge responses on the same patient in the home based care grouping. This was overcome with Excel in the same way that the BETA data was prepared. Therefore the final dataset for analysis had 634 observable single raw scores of 634 persons.

This raw score dataset which was used for the first GAMMA analyses run across the 8 items. The next step was to implement [Linacre's \(2010\)](#) advisory that clinical observations with under fitting responses over 1.7 mean square logits are usually associated with careless mistakes too unpredictable for Rasch measurement development. It should be removed for calibration. Therefore the most miss-fitting data (< 1.7 MNSQ logits) were removed leaving the remaining data set of 570 responses free of under fitting data and also theoretically free of local dependence. This raw data were used for the Rasch analysis of the GAMMA.

7.5.3 Rasch calibration

The first step was to check for disordering of the items by running the category probability curves of the eight items. It produced unsatisfactory results as disordering of categories were observed across all the items. From the probability curves it became clear that the nurses have difficulty in observing seven different levels of independent living. However, the Rasch analysis revealed in which categories nurses have problems with distinguishing between, and suggest collapsing with neighbouring categories. When this was

done, the ordering of categories improved significantly. This iteration process eventually revealed that the GAMMA functions optimally as a linear interval scale with a four category structure across all eight items, rather than a seven category structure as originally intended.

The next step was to verify if the GAMMA raw scores comply with the four basic requirements, explained by Iramaneerat et al (2008), to make valid inferences from the Rasch analysis on the raw data. These are local independence of the test data, uni-dimensionality of the items, monotonicity of the latent trait and finally non-intersecting probability curves. All these assessments rendered positive results. With the admission criteria to the Rasch analysis thus satisfied, the data was ready for the measure and calibration distribution matches.

Figure 7.2 shows how well the researcher managed to construct the person ability and item difficulty on a common logit scale represented as a straight vertical line. The interpretation was discussed in Chapter Four. In **Figure 7.2** there is an acceptable person ability and item difficulty match, but both a ceiling and floor effect is present. This can be explained that the sample selection does not fully fit the anticipated range of the scale. Firstly, persons living independently in selected retirement villages were scored. This includes numerous newly retired persons being fully independent. No selection criteria were used to select an appropriate sample for the range of the scale, e.g. persons more than 75 years old. Secondly, a floor effect was noticed because a substantial number of home based care patients were included in the database and they were scored whilst convalescing from acute care. This made them incapable of participating in any of the independent living activities at the time when scored.

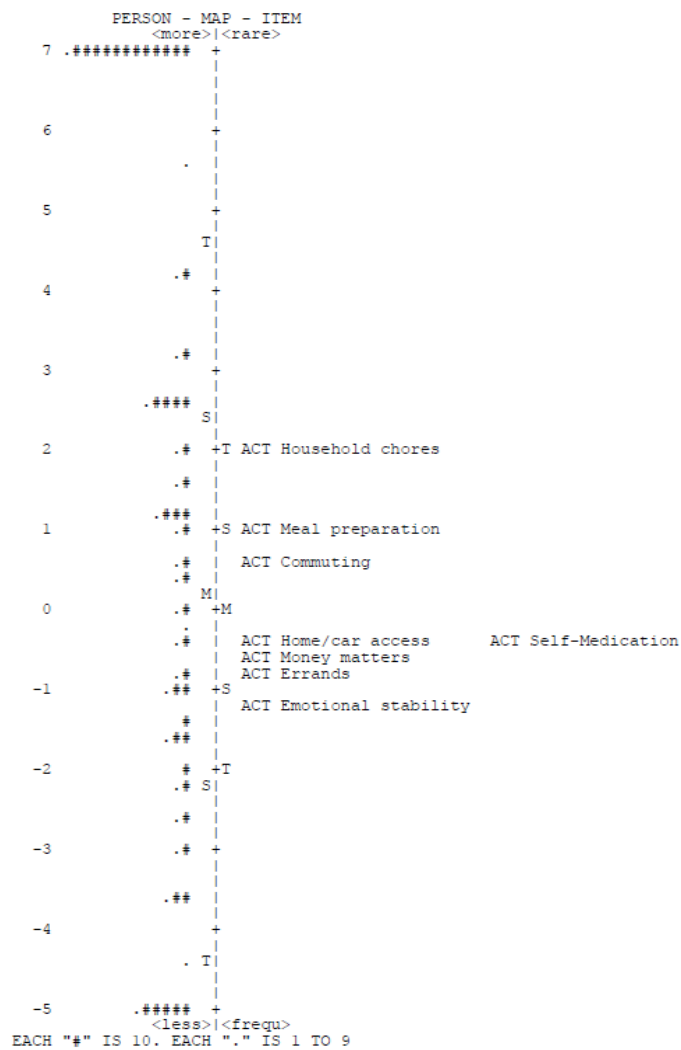


Figure 7.2: Variable map of patient ability and item difficulty in the GAMMA scale.

7.5.4 Results on category functioning

The overall results of the analysis on the functioning of the categories can be seen in [Table 7.1](#) and [Table 7.2](#). The data in the “New structure” column in [Table 7.1](#) must be interpreted the same as explained in the BETA category functioning section. According to [Linacre’s \(2004\)](#) suggested guideline estimates when assessing quality in category functioning, the GAMMA performed as follows:

- All the GAMMA items are orientated with the latent trait variable. [Table 7.2](#) shows all the items in the Point-Measure Correlation (PT MSE CORR) column with high positive values (>0.80). Linacre warns against negative coefficient values, or positive values lower than 0.30.
- A minimum of 10 observations is required in each rating category and the GAMMA sample fulfils that guideline as can be seen in [Table 7.1](#).
- The observations are regularly distributed across all rating categories. The frequency distribution loadings (OBS COUNTS) on the categories can be considered very good on items 1,2,3 and 4 and good on items 5,6,7 and 8.
- As required the average measures are advancing monotonically with category as the observed average (OBSVD AVRGE) indicates on [Table 7.1](#).
- The outfit mean square (Outfit MNSQ) values for the categories are less than 2.0 as is evident on [Table 7.1](#). A single value marginally overstepped the recommended guideline, but this was considered insignificant and not warranting remedial work to force these values lower during this level of analysis.
- The thresholds advance orderly with categories as seen in the column Structure Calibration. These thresholds correspond with the intersecting points between the CPCs in [Figure 7.3](#). It can thus be assumed that the GAMMA categories take increasing levels of the latent trait to be observed in higher categories.
- However, there is good news for Linacre's 8th and final guideline as none of the GAMMA thresholds is exceeding the 5 logit margins, except for item 2 which shows an insignificant indiscretion of 5.02 logits.

The newly structured GAMMA still have 8 items, but each now has 4 categories and 3 thresholds, totalling 24 new GAMMA thresholds. Of all the new thresholds, 21 advances by at least 1.0 logits in the Structured Calibration column, indicating that these neighbouring categories are performing within range as suggested by [Linacre \(2004\)](#), and are clearly separable and functioning independently. However, of the three

underperforming categories one is in the marginal range (item 5: Errands advancing with 0,80 logits), one is outside the marginal range (item 7: Self Medication advancing with 0.64 logits) and one is in the unacceptable range for measurement (item 6: Money Matters advancing 0.21 logits).

Table 7.1: Results on the GAMMA category functioning.

| Item | Category Label | New structure | OBSVD COUNT | OBSVD AVRGE | OUTFIT MNSQ | Structure Calibration |
|-----------------------|-----------------------|----------------------|--------------------|--------------------|--------------------|------------------------------|
| 1.Meal preparation | 1 | | 136 | -2.47 | 1.17 | None |
| | 2 | 1222334 | 148 | -0.45 | 0.88 | -3.02 |
| | 3 | | 133 | 1.66 | 0.81 | -0.15 |
| | 4 | | 153 | 3.54 | 0.94 | 3.17 |
| 2.Household chores | 1 | | 141 | -2.38 | 1.13 | None |
| | 2 | 1222334 | 174 | -0.13 | 0.99 | -3.94 |
| | 3 | | 125 | 2.12 | 0.88 | -0.54 |
| | 4 | | 130 | 3.29 | 1.24 | 4.48 |
| 3.Home access | 1 | | 111 | -3.06 | 0.67 | None |
| | 2 | 1222334 | 92 | -1.13 | 0.87 | -2.24 |
| | 3 | | 163 | 0.98 | 1.12 | -0.55 |
| | 4 | | 204 | 2.17 | 1.56 | 2.79 |
| 4.Commuting | 1 | | 148 | -2.52 | 0.81 | None |
| | 2 | 1222334 | 128 | -0.50 | 0.82 | -2.32 |
| | 3 | | 108 | 1.55 | 0.67 | 0.14 |
| | 4 | | 186 | 3.04 | 0.91 | 2.18 |
| 5.Errands | 1 | | 123 | -2.92 | 0.86 | None |
| | 2 | 1223344 | 77 | -1.19 | 1.06 | -1.30 |
| | 3 | | 73 | -0.29 | 0.42 | 0.25 |
| | 4 | | 297 | 2.13 | 0.62 | 1.05 |
| 6.Money Matters | 1 | | 115 | -3.06 | 0.70 | None |
| | 2 | 1222234 | 126 | -1.07 | 0.46 | -2.14 |
| | 3 | | 64 | 0.75 | 0.56 | 0.96 |
| | 4 | | 265 | 2.32 | 0.96 | 1.17 |
| 7.Self-medication | 1 | | 143 | -2.65 | 0.81 | None |
| | 2 | 1222334 | 62 | -1.24 | 0.49 | -1.01 |
| | 3 | | 97 | 0.41 | 0.50 | -0.37 |
| | 4 | | 268 | 2.28 | 0.89 | 1.38 |
| 8.Emotional stability | 1 | | 86 | -2.03 | 2.00 | None |
| | 2 | 1122334 | 90 | -1.36 | 1.71 | -2.33 |
| | 3 | | 138 | -0.29 | 1.61 | -0.11 |
| | 4 | | 256 | 2.13 | 1.38 | 2.44 |

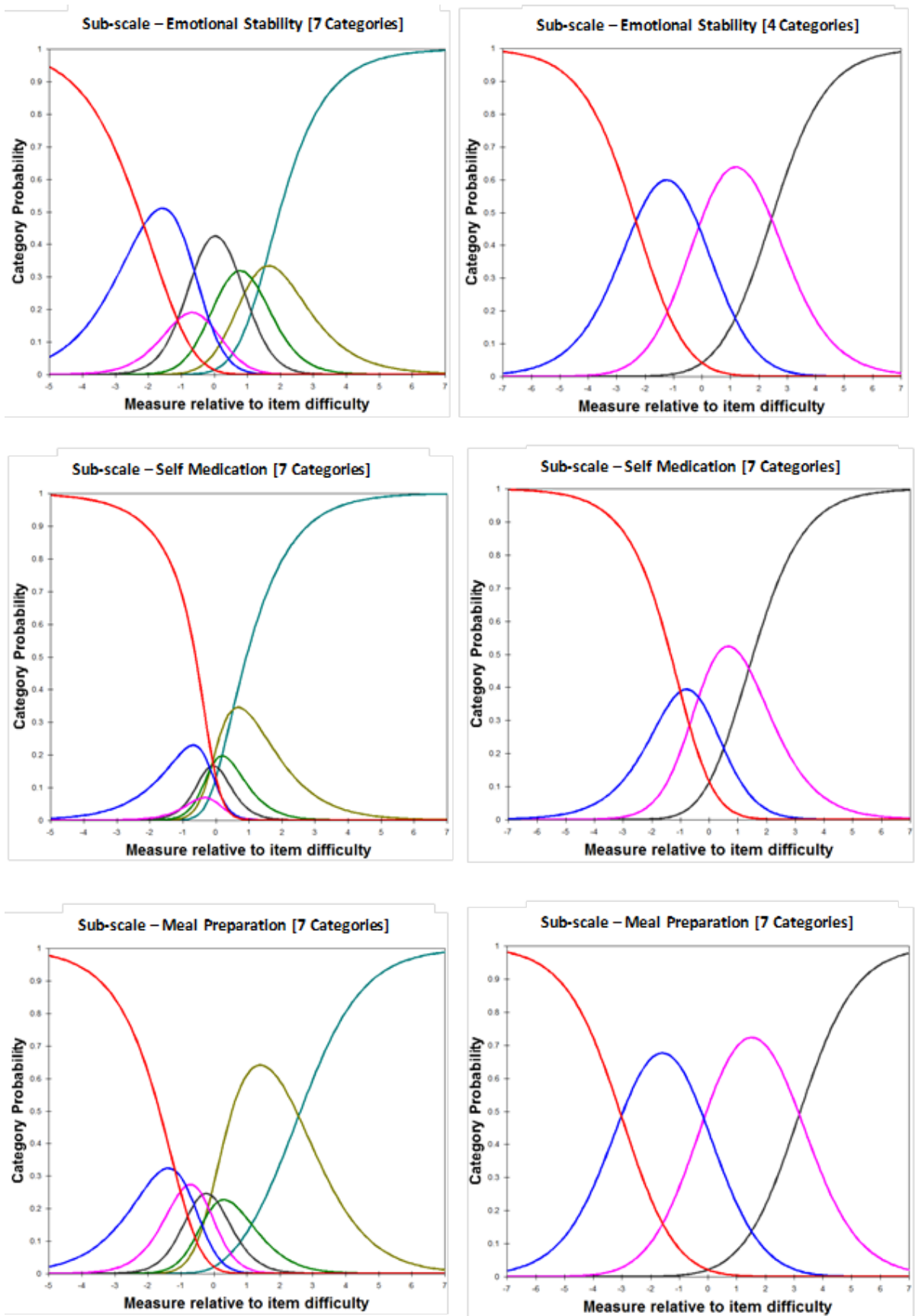


Figure 7.3: Examples of the GAMMA category probability curves before and after collapsing.

Nevertheless, [Linacre \(2004\)](#) advises that these 8 guidelines should be evaluated in conjunction with the Category Probability Curves (CPC). These curves indicate how the category response structure is predicted to work for any future sample, provided it worked satisfactorily for this sample. For a lack of space and to prevent repetition, the CPC of only three items are presented in [Figure 7.3](#).

After the remedial work has been done, the peaks of the new categories presented in [Figure 7.3](#) are all in ascending order along the latent variable of each item. At some defined point on the latent variable; each category in turn is the most probable than any other one of the categories. Furthermore, the cross-over points between the categories are ordered; e.g. the descending curve of each category clearly crosses the ascending curve of the neighbouring category. These cross-over markers are the equal probability points or the thresholds or the parameters of the CPC. This investigation into all the new category structures of all the items concluded that it was ordered and no further remedial action for the categories was required.

7.5.5 Results on item functioning

The next step was to study the functioning of the GAMMA items. The Rasch model indices selected for reporting on the fit statistics for items were both the INFIT and OUTFIT MNSQ values, the point measure correlation (PT MSE CORR), Rasch reliability for person and item, and the variance experienced by measure ([Table 7.2](#)). According to the literature these parameters are the most widely referred to and commonly used in item fit statistics.

The INFIT and OUTFIT MNSQ values are the core statistics to verify if the scale fit with the Rasch model or not. It also indicates how closely the scale appropriates the Rasch model. When values are around 1 logit, the measure is considered accurate. However, for clinical scales such as the GAMMA sub-

scales, [Linacre \(2010\)](#) suggests INFIT and OUTFIT MNSQ value ranges between 0,5 and 1,7 as reasonable for quality measurement. [Fischer's \(2007\)](#) criteria have more specific ranges: Good (0.5 - 2.0) Very good (0.71 – 1.4) and Excellent (0.77 – 1.3). The fit statistics of the first four items can thus be considered as being “excellent”, items 5, 6 and 7 as being “very good” and item 8 as being “good”.

As mentioned above, the PT MSE CORR shows high positive values (>0.30) indicating the items are all orientated with the latent trait variable.

The Rasch reliability on items and persons achieved values well into [Fisher's \(2007\)](#) “excellence” criteria (> 0.94). Finally, the results in the Variance Explained by Measure column of [Table 7.2](#) is rated by Fischer as being in-between “good” (60% - 70%) to “very good” (70% - 80%) rating scale qualities.

Table 7.2: Results on the GAMMA item functioning.

| Items | Sample | Categories per item | Outfit MNSQ | Outfit MNSQ | PT MSE CORR | Rasch RELIABILITY Person/Item | Variance explained by measure Emp / Mod |
|-------|--------|---------------------|-------------|-------------|-------------|-------------------------------|-----------------------------------------|
| 1 | | 4 | 0.98 | 0.94 | 0.91 | | |
| 2 | | 4 | 1.03 | 1.02 | 0.92 | | |
| 3 | | 4 | 1.13 | 1.09 | 0.86 | | |
| 4 | | 4 | 0.79 | 0.80 | 0.90 | | |
| 5 | 570 | 4 | 0.79 | 0.70 | 0.83 | 0.99 / -0.99 | 70.2% / 69.9% |
| 6 | | 4 | 0.75 | 0.63 | 0.86 | | |
| 7 | | 4 | 0.79 | 0.64 | 0.85 | | |
| 8 | | 4 | 1.80 | 1.63 | 0.80 | | |

7.6 Conclusion

From the raw score data, the Rasch concluded that nurses have difficulty in

distinguishing between some categories as there is disorder in the category function across all the items. This meant either the category definitions were too vague or the distinctions between the patient abilities are too difficult to be observed. However the Rasch diagnostic process indicated that if one collapses the original 7 category structure into a four category structure, the data not only makes more measurement sense, but also fit the Rasch model. The Winstep software provided the stepwise diagnostic method for a structured collapsing of the categories to secure the best available measurement outcome. The final result can be seen in [Table 7.1](#).

In the final analysis the GAMMA satisfies the Rasch Model with a “good” fit according to the [Fischer \(2007\)](#) criteria. The first four items provided “excellent” fit statistics on both the category analysis and the item analysis. Items 5, 6 and 7 falls within Fischer’s the “very good” measurement fit grouping and item 8 is considered an example of “good” fit. The nursing logic why different items arrived at different measurement qualities might be as follows:

- To do the observations on errands (item 5) and money matters (item 6) the nurses have to rely on secondary raters to function as proxy observers. This increased the difficulty of arriving at scores and could explain the decrease in accuracy of the observations as reported by the Rasch statistics.
- In the case of self-medication (item 7) the nurses are faced with a legal situation whereby they are not allowed to provide patients with daily or weekly dispenser pill boxes once they have become aware that the person is defaulting. In such instances the nurses must physically administer the medication to the person at each dosage. This situation makes it difficult to establish the person’s ability when the person is not allowed to perform at their ability.
- The observation of emotional stability (item 8) may be improved with the provision of clear definitions and training on the development of emotional stability during the aging process. Nurses may have a difficulty in distinguishing between the behaviours of persons with an

existing hostile personality, and the symptoms associated with agitation, frustration, anxiety and mood swings as a result of functional decline.

The Rasch outcome thus created a new awareness on how the GAMMA items and categories function, and identified the weak points in the structure. In retrospect the researcher should have foreseen these fault lines, but with scientific proof now available that these faults do exist, and even pointing to the faulty areas, more focused corrections can be made.

In conclusion, the GAMMA has significant potential to be implemented as a routine nursing measure. It is suggested that some remedial changes must be made to the observational format to improve accuracy in the observations of items 5-8 and then repeat the data collection and the Rasch analysis.

CHAPTER EIGHT: THE DELTA

What we really need now in mental healthcare is quality of improvement (QI) measures providing practice-based evidence (PBE) data as it transpires in our real-world practices. It seems that evidence-based practices (EBP) data to model scientifically sound practices are not effective to establish performance and outcomes.

Dr David J Hellerstein 2009

8.1 Introduction

Selecting between patient outcomes measures in the field of acute mental healthcare appears to be a complicated process. Most of the existing severity measures aim at a single impairment and do not provide broad severity data across all impairments. Decision-makers must therefore rely on additional inferences for economic and outcome studies and subsequent policy decisions. Indeed, [Stant et al. \(2007\)](#) warned that this practice of generalisation has considerable margins of error and that all such inferences must be treated with caution. Ideally, decision makers require data provided by a generalised measure across all impairment groups. More so if such a measure could provide routine longitudinal patient outcomes data. Such information, if validated, would unlock research opportunities for outcomes analysis and subsequent reliable economic and policy inferences.

Scales currently competing to comply with these requirements are structured self-report questionnaires such as the Behavioural and Symptoms Identification (BASIS) designed by [Eisen and Dickey in 1996](#). However, as a subjective questionnaire it has limited use in a country such as South Africa, where although rich in cultural diversity, levels of illiteracy remain significantly high. A strong competitor, according to [Meagher, O'Brian, Pulella and Brosman \(2009\)](#) is the Health of the Nation Outcomes Scale (HoNOS), widely used in Britain,

Australasia, Canada and some European countries, although there is little consensus in the literature to its usefulness as a routine measure and its accuracy to be used as an outcomes measure. In fact, [Bebbington *et al* \(1999\)](#) and [Adams, Palmer, O'Brian and Crook \(2000\)](#) argue against the use of measures such as HoNOS as a standardised routine measure, as it has been found to have questionable validity and only a tenuous relationship with patient severity. [Lakeman \(2004\)](#) took a strong nursing perspective that clinician-rated standardised tools such as the HoNOS have little if anything to do with, or to offer towards the service users' recovery process. [Trauer, Callaly and Herman \(2009\)](#) reported that the HoNOS enjoys limited acceptance by healthcare professionals as being a useful measure. [Meagher *et al* \(2009\)](#) cautioned that service needs cannot be judged merely upon a cross-sectional assessment of active symptomatology on admission and discharge, as rendered by the HoNOS, but rather require routinely observed and recorded longitudinal data on patient changes as they occur.

A different competitor is the Global Assessment Functioning (GAF) scale frequently used by psychiatrists and often referenced in research. It constitutes the fifth axis of the Diagnostic and Statistical Manual of Mental Disorders (DSM), currently under its fifth review. The GAF is not a measure, but rather a clinical classification scale based on the clinician's judgment of the subject's overall level of functioning ([APA 1994](#)). The literature is expressing significant concern about the GAF's subjectivity, its lack of detail in its user guidelines and poor validation ratings ([Aas 2010](#)). In an attempt to satisfy these shortfalls, researchers and developers are currently trying to improve, adjust or repair the inadequacies of the GAF ([Aas, 2011](#)).

Even with all the pressure on mental healthcare workers to find a solution, [Salvi, Leese and Slade \(2005\)](#) reported little consensus on which outcome measures to use in mental healthcare. They concluded that meaningful and comprehensive clinical information could only be provided by a combination of existing measures. However, [Aas \(2010\)](#) cautioned that if the number of scales is increased, there may be a longer learning time for the scoring method,

scoring becomes more time consuming and less easy to use, and the outcomes analysis of the data might become more complex.

Lately, remarkable pressure is mounting internally from clinicians challenging the randomised control studies of evidence-based practice (EBP) as too far removed from the real world of mental healthcare practice. Clinicians are advocating a parallel consideration for practice-based evidence (PBE) as being more connected to the context of real practice. They rate firsthand knowledge and experience of what works, what needs to change and how it may change as experienced by their patient-based outcomes higher than prescriptive formularies of academics ([Hellerstein 2008](#); [Warrol 2007](#)). Irrespective of the ongoing debate between EBP and PBE preferences, the point of consensual departure still requires the need for routine monitoring of patient progress. ([APA Presidential Task Force 2006](#))

The next question is: which profession is best suited, placed and skilled to monitor patient progress routinely? Although [Meagher et al. \(2009\)](#) reiterated consensus that a multidisciplinary team is the preferred approach for mental health services to record the complexities of severe mental illness, the authors argued that nurses' may have an advantage over their peers. Their continuity and proximity as primary caregivers enable direct observations of changes as they occur. It is well recorded that this advantage create a rich intuitive knowledge integral to the nursing practice, and [Billay et al. \(2007\)](#) emphasised it must be recognised as a legitimate form of nursing skill. In their line of duty, nurses already report their observations descriptively on patient severity and progress. All these factors indicate the nursing profession, however, a literature review revealed no validated nursing framework whereby mental healthcare nurses could score their patients based on their objective observations. The purpose of this study is to introduce and validate such a new rating framework for the nursing profession.

8.2 Development of the DELTA

In 2007 the South African Database for Functional Medicine (SADFM) conceptualised, designed and developed a nursing scale by exploring the intuitive knowledge of nurse practitioners working in acute mental healthcare settings. The aim was to provide the mental healthcare nurse with a standardised framework to observe and score the severity of an acutely ill mental health patient. The scale was designed to be used routinely by registered mental healthcare nurses. It was named the DELTA, being the fourth in an interconnected family of SADFM nursing measures across the continuum of nursing practice.

The DELTA's design and development was part of a qualitative study whereby the developer interviewed nurses individually and in focus groups over a period of six months until data saturation was reached. The developer explored the nurses' experience of what enquiry or observation prompted them to conclude a nursing diagnosis that a mental healthcare patient is gravely acute, irrespective of the impairment. The consensus on this information was used as the lowest score for the DELTA. From this lowest score, nurses readily explained what line of enquiry prompted them to recognise early improvement e.g. a turnaround in health status. Consensus on this information formed the second lowest score of the DELTA. The follow-up scores were achieved using the same methodology of enquiry until saturation was reached at seven clearly distinguishable scores of severity. The seven severity scores became the seven categories of the DELTA. These DELTA categories were not linked to any diagnoses, they were linked to nursing observations that conclude the levels of severity of mental illness. These seven levels and the observational methodology to distinguish between them, became the basic information required to develop a decision tree for the novice mental healthcare nurse to arrive objectively at the correct category based on nursing observations.

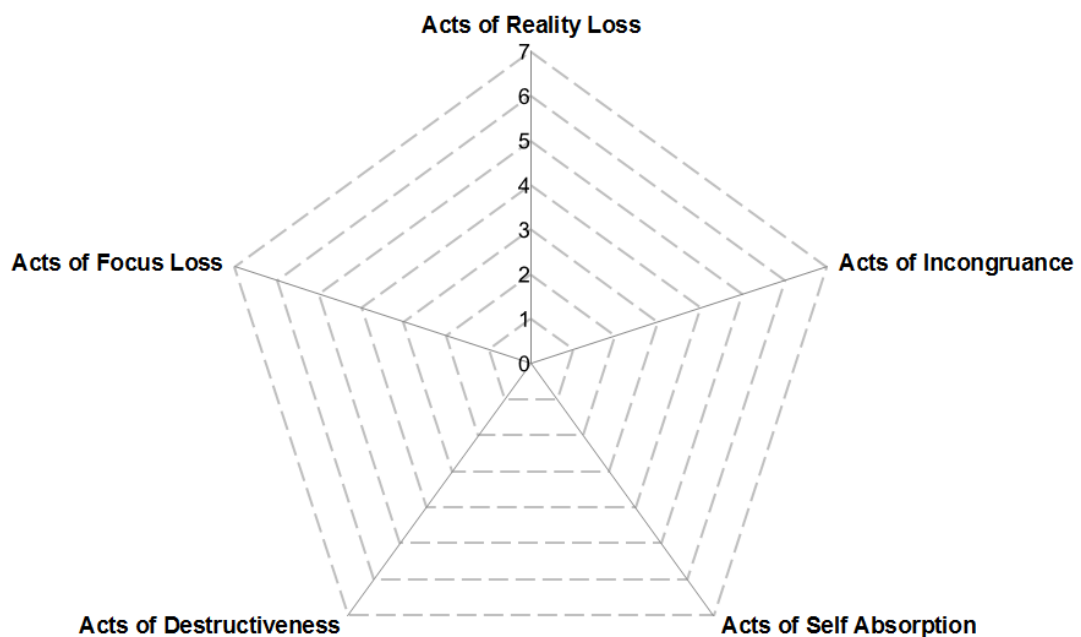
The next step was to identify the latent variables associated with the major symptoms, signs and behaviours. The respondents identified five major latent

variables, each with its own generic groupings based on observable functional loss. These five latent variables became the DELTA items, each with its own unique cluster of psychiatric terms and definitions underpinning it, as illustrated in **Table 8.1:**

Table 8.1: Description of DELTA items.

| Item | Symptoms, signs and behaviours |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Acts of Reality Loss | Reality gaps refer to disrupted thought processes that manifest in the Subject not being able to understand the reality. This also includes delusional comprehension where the Subject sees, hears and believes “things” that are clearly not present or not true (<i>e.g. delusions (inappropriate to culture or a fixed false belief), hallucinations (false perception with no stimulus and could be audible, visual, tactile, olfactory, gustatory), formal thought disorder, decreased meaning in life, impaired (poor) insight, guilt, self blame, blaming others, conspiracy theories etc.</i>). |
| Acts of Incongruence | Incongruous acts or behaviour refer to any verbal or non-verbal expressive reactions of the Subject that appear inappropriate because they are very different from the surroundings, or are not suited to the situation (e.g. emotional lability, fatuousness, mood swings, incongruent emotions, neologisms, denial, bluntness, apathy, phobias, muscular spasms secondary to anxiety or learnt behaviour, defence mechanisms, withdrawal, projection, pain of non-pathological origin, agitation, psychosomatic disorders, anxiety and panic attacks). |
| Acts of Self-absorption | Self-absorption refers to any act or behaviour that indicates that the Subject is so pre-occupied with the self that s/he has difficulty considering other people (e.g. obsessional behaviour, obsessional thinking, compulsion, kleptomania, asocial behaviour, preoccupation with fantasies, perseveration, cravings, intrusiveness, tantrums, rage, violence, aggression, allurements, seductive behaviour, narcissism, provocation, verbal provocation, low level of motivation towards socially negotiated or culturally prescribed behaviour, low level of volition, manipulation, immediate gratification, childishness, poor social judgement). |
| Acts of Destructiveness | Destructiveness refers to any act or behaviour (short or long standing), which most probably results from an inability to resolve problems (e.g. verbal abusiveness, emotional abusiveness, destructiveness against people, animals, plants or objects, self-mutilation, self-neglect, dietary disorders, substance abuse, procrastination, hypersomnia, wilful stealing, emotional dependency, low frustration tolerance, and occasional suicidal thoughts). |
| Acts of Focus Loss | Concentration gaps (blank attacks) refer to spells of loss / lack of focus, lack of memory, blank periods, absentmindedness, thought block, loitering, wondering, disorientated, distraction (e.g. due to insomnia, stress mismanagement or PTSD, exposure to trauma, fatigue, flashbacks, worry, anxiety, etc.). |

The DELTA framework thus consists of a domain (“severity of mental illness”) with five items each having seven categories of severity as illustrated in *Figure 1: Delta Profile*. The Delta structure represents a framework of ordered qualitative measures estimated or obtained directly from a nursing observation; it is not derived from other measures. From clinical experience it is known that the intuitive distance between categories might not be the same, for example, when one considers the item ‘reality loss’, to progress from category 2 to category 3 might need more skill than to progress from category 4 to category 5 (see [Figure 8.1](#)). In other words, the categories of an item are of an ordinal measurement nature. By using the Rasch measurement model, the ordinal scores are transformed into linear interval measures. Thus, to operationalise the DELTA as a fundamental measure for analytic and reporting purposes, the study requires DELTA data to undergo a process of concatenation so that the categories can be seen as equal to allow addition ([Luce & Tukey 1964](#)).



[Figure 8.1](#): Polar graph representing the Delta scale design.

To collect DELTA data, the respondents collaborated in the development of a user manual for training, testing and implementing the DELTA into the nursing process of a 116-bed state facility where, for two years, DELTA admission, intermediate and discharge scores were done on all acute mental healthcare patients admitted (>18 years) into the acute facility.

8.3 Conceptual framework of the DELTA

The Delta measure is introduced to be a new mental healthcare nursing measure to be used on adult persons with acute mental healthcare illness receiving inpatient care. It is intended to render objective cross-cutting measurements on the severity of the mental illness irrespective of the diagnosis or underlying pathology. Structurally, the Delta is a short measure with five items, each having a hierarchical algorithm arriving at seven categories. Thus the total summed raw score range is between 5 and 35 with the lower scores indicating the severity of the illness. No patient self reporting is required and its objective framework of observation and scoring is anticipated to be familiar to mental healthcare nurses. Nurses need to undergo training and testing for proficiency before administering the Delta measure. The training is a once-off event that lasts 4-6 hours.

8.4 Qualitative study: nursing utility

8.4.1 Sampling and data collection

For the qualitative analysis, a focus group consisting of six nurses provided the descriptive data. Five were registered mental healthcare nurses, with one having a master's qualification in mental healthcare nursing. The facility operations manager had no formal mental healthcare qualifications,

but substantial experience in managing nursing services in this specific acute mental healthcare facility. They were all working in the acute mental healthcare section and all had formal and in-service training on the DELTA. They used the DELTA routinely in their work place. Each had between seven months and two years experience in using the DELTA.

The four broad questions selected for discussion were regarding: the DELTA's overall usefulness to the nursing profession; the ease to embed it into the nursing process as a routine measure, its ability to become a universal language as measure of severity of mental illness, and its potential to improve quality of nursing care. The responses were recorded and transcribed.

8.4.2 Data analysis

Similar to the explanation in Chapter Four, the DELTA also theoretically had nursing utility as every step during the design and development of the DELTA was tested with the same four questions described in Chapter Four. Therefore these four questions, with previous consensus, provided some prediction about the variables of interest, and therefore predetermined the initial coding scheme to anchor the categories of analysis (Hsieh & Shannon 2005). The analysis thus followed a similar deductive content analysis approach (Elo & Kyngras 2007).

8.4.3 Results from the professional nurses' focus group

The results are discussed under the five categories identified. The references in brackets refer to the page and line numbers in the transcriptions where the evidence was found.

- Usefulness

There was consensus that the DELTA embodies a useful *“new mindset”* of measurement to the mental healthcare nurse. However, for the novice with little or no previous mental healthcare experience this new approach was initially difficult to surmount, but *“within months”* the DELTA became *“very easy to use”* (p6:12-14). Once mastered, the nurses reported new insights in practice as they found themselves able not only to quantifying *“if the patient is improving or not improving”* (p2:9); but also to evaluate and record *“how the patient is improving from admission to discharge”* (p3:16-22). This ability has empowered them to guide the team as nursing DELTA scores became *“a good indicator to see if things start to go wrong for the patient”* (p3:24-29). As they started to use it routinely they also found themselves to be *“very successful to further evaluate deterioration or improvement in patient scores and even adjust treatment according to the scores”* (p5: 1-3). Overall the nurses reported a new skill whereby they found *“it very helpful to be able to evaluate if the patient is improving or not improving”* (p2:9). Additional content validity inferences from the data confirms that the DELTA scores correlate with nurses’ clinical

judgment, and that they have developed a confidence in the DELTA scores to the extent that they advocate adjusting treatment and services based on their recordings (p3: 11-16).

- Nursing process

All responses indicate that the nurses use the DELTA routinely during the assessment, diagnosis, implementation and evaluation phases of the nursing process and found it to be *“beneficial to the nursing process”* (p8: 3-7). The nurses provide the evidence how the DELTA *“mindset”* helped their nursing process as they followed its nursing logic. They reported that *“previously it was difficult to interview, but if you know the content of the DELTA it makes it now much easier to interview as we now know what to look out for* (p7: 27-29). It was also reported that *“it is now easier to make a nursing diagnosis and establish the severity of the diagnosis”* (p8: 9-10).

The nurses highlighted four awareness features of the DELTA which had become routine embedded into their nursing process: firstly, the DELTA requires nurses to identify specific symptoms of the patient which direct them to the appropriate nursing diagnosis (p4: 5-8; p8: 8-9); followed by the rating of the severity which led them to devise and implement appropriate nursing care plans (p4: 5-8); and the DELTA's ability to raise warnings on suicidal risks (p5: 16-17). Finally, longitudinal scoring and recording of patient severity are used by nurses to advocate for treatment changes at the multi-disciplinary team meetings.

- Uniform language

As the DELTA scores become routinely embedded into the nursing process, the changes in the scores become helpful references of communication between nurses and doctors to discuss diagnosis and treatment plans (p2: 24-26). Frequent discussions between nurses and therapists are also evident on the issue of actual versus potential severity scores. Nurses claim they score the actual performance of the patient, as they have the benefit of observing the patient for 24 hour per day. However, the nurses also claim that they reach consensus with the therapists at the multidisciplinary meetings (p3: 1-6).

Additional categories were identified under universal language: Nurses are promoting the DELTA data as valid nursing observations because they are actual observations taken over 24 hours. Substantial evidence was found that the multi-disciplinary teams may have become reliant on the nursing DELTA scores as the preferred outcomes measure to evaluate patient progress or decline. Nurses report that multidisciplinary teams base their considerations to adjust treatment plans on their nursing scores; and described this acknowledgement to their contribution as "*exciting*" (p4: 28-29; p51-9). Nurses further argue their ability to score a patient as a suicide risk and use this score as motivation to management to supply the patient with one-to-one nursing care. This phenomenon was mentioned to indicate how the DELTA supports the nursing process to improve the conditions of care (p5: 25-28)

- Quality of nursing

Whilst probing the issue of nursing quality care, there was consensus that the DELTA “*definitely*” raised nursing awareness of the patient symptoms, severity and needs. As the DELTA scores became the universal language at the multidisciplinary meetings (p5: 15-16, 20-22), and as the nurses had to defend their scores, the nurses’ interaction skills with the patient and their interviewing skills and observations of patient behaviour has improved. Being more aware and knowing the patient better also results “*in rendering better conditions of care*” (p6:2). Thus, the DELTA being a nursing measure and producing scores to calculate patient outcomes not only “*contribute to improving the quality of nursing care*” (p6:4-6); but also contribute to “*improving the nursing skills*” (p6:8).

There was strong evidence that nurses regularly discuss DELTA scores amongst themselves to evaluate their nursing process. A decline in the scores indicates a regression that necessitates a relook at the nursing interventions and care-plans, and a gain indicates improvement that may or may not require adjustments to treatment. A respondent described these discussions as an “*exciting*” experience (p5: 8-9). Upon discharge, the difference between the admission and discharge scores are calculated and presented in a graph to indicate the improvement as an expression of the patient outcome. Respondents referred to this graphic outcome as “*good to see*” (p8:21-25)

- Perceived problems

The nurses reported uncertainty amongst themselves when they reached a one-score-difference e.g. 3 or 4, on the same day and on the same patient, not knowing which one to record for that day. They were asking for clarification and the researcher suggested recording the lowest score when in doubt.

Another uncertainty raised was the ideal score achieved to be discharged from acute care. As patients are discharged from the acute mental healthcare facility when they are able to function in the community with support from the out-patient clinics, the actual discharge scores will vary between 5, 6 and 7

depending on the item, diagnosis and community support available, and not on the potential scores of 7. The researcher suggests further clarification in the training manual.

The final perceived shortcoming is the descriptions of definitions used in the DELTA. The descriptions are not readily understood by non-registered mental healthcare nurses and require further simplification to make the DELTA accessible to skilled and experienced registered nurses, but without formal training in mental healthcare nursing. A decision must be considered whether the DELTA must be used by non-registered mental healthcare users.

8.5 Quantitative study: construct validity

8.5.1 Sample and data collection

DELTA data were collected for the quantitative analysis on 1995 adult patients over the age of 18 years, admitted into a 116-bed mental healthcare care hospital. The hospital is managed by the South African government as a regional specialist mental healthcare facility. At the onset it is important to mention the state's admission and discharge policies for regional mental healthcare hospitals. This stipulates that all patients requiring acute mental health services must first be admitted into one of the four acute hospitals in the region for a maximum period of 72 hours to be stabilised. Only after this period can they be transferred and admitted into the region's specialist mental healthcare hospital where the DELTA data was collected. The effect of this policy is that acutely psychotic or suicidal patients are rarely observed in the regional hospital. Furthermore, the discharge policy is to discharge patients to community clinics and infrastructure as soon as mental stability has been achieved. In other words, the patient has not improved 100% (e.g. did not achieved the highest DELTA ratings) when discharge to community services

are affected. The effect of these policies is that disproportionate numbers of patients with very low and very high DELTA scores are expected to be observed.

The data were collected over a period of two years. All patients admitted into the facility were included in the study, irrespective of their diagnosis or underlying pathology. No exclusion criteria based on gender, race or ethnicity prevailed. All admission, weekly intermediate and discharge DELTA observations were recorded, totalling 9 890 raw DELTA scores.

All professional registered nurses working in the pilot units were provided with the Delta manual and trained to use the scale. The nursing team designed and developed their own nursing process documentation to record the DELTA raw scores. All admissions were recorded within 48 hours. Intermediate score changes were recorded as they are observed and were presented as a nursing progress report at weekly multi-disciplinary meetings. All discharge scores were recorded on the day of discharge. An electronic web-based application was provided to import the admission, intermediate and discharge scores from the nursing documentation.

8.5.2 Data preparation

The original 9413 raw scores collected from a sample of 1955 patients, included their admission, intermediate and discharge scores. While it is feasible to include all score levels that represents the full range of the latent variables, local dependency becomes a concern when scores of the same patient on admission, intermediate and discharge dates are included. To overcome local dependency, a random sample was created by using an Excel random number generator. This function assigns a random number within a specified range using a uniform distribution to each score entry. Thus a random sample of each of the admission, discharge and intermediate group scores was created, identified and added. Any duplicate patients were also then eliminated by

running these patients through the Excel random number generator again. It was expected that the residual dataset with 1379 responses would be local dependency free.

The next step in the data preparation was to remove under fitting responses. According to [Linacre \(2010\)](#) the reasonable item mean-square statistics for all clinical observations should not exceed mean-square values of 1.7. Values of statistical fit statistics greater than 1.7 are termed “under fit”, meaning that the responses are too unpredictable from the Rasch model’s perspective, and make the calibration inaccurate. “Under fit” responses are mostly caused by novice raters making obvious errors because they are either guessing or not applying their minds fully when arriving at a score. These careless mistakes create unexpected outliers in the data and the Rasch model identifies these unpredictable response patterns. Therefore all under fitting observations greater than 1.7 were removed leaving a raw score database of 1152 responses on which to proceed with the RMM.

8.5.3 Rasch calibration

The first step in the RMM calibration was a preliminary testing on the data to verify if the DELTA’s raw scores are suitable for the Rasch analysis. The four basic assumptions according to [Iramaneerat et al \(2008\)](#) are:

- *Local dependency* which was addressed in the data preparation above.
- *Unidimensionality*: The relative fit parameter for unidimensionality tested was the Eigen value that should be < 2 (Linacre 2010). The DELTA value reported as the unexplained variance in the first contrast was 1.6 (5,3%) indicating that no secondary dimension of concern is observed in the 1152 data set.

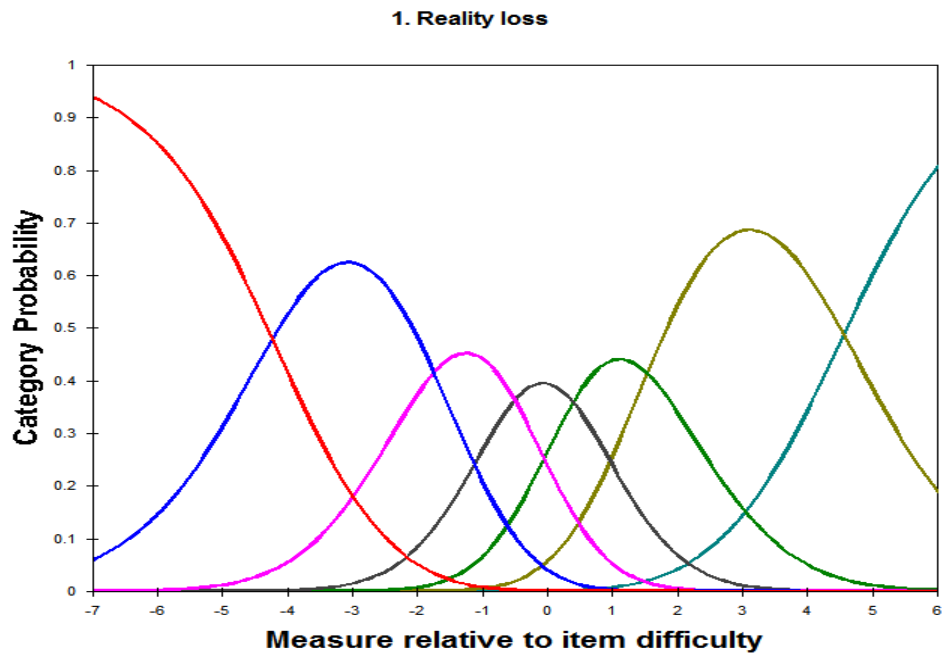


Figure 8.2: The Category Probability Curve of Item 1: Reality Loss.

- Monotonicity:* Winsteps analysed an observed average logit measures and reported that it advanced orderly with rating scale categories, meaning the DELTA has satisfied monotonicity. This can be seen in the observed average (OBSVD AVRGE) column in Table 8.2 where no disordering of the DELTA categories can be observed. This means the probability that a patient will comply with an item difficulty is monotonically increasing over the range of the latent trait (Sitjtsma & Molenaar 2002)
- Invariance:* A patient's ability range should correlate with the scale difficulty range. A patient with low ability should respond to low difficulty categories and a patient with high ability should respond to high difficulty categories on any item of the scale. Winsteps produce graphs with Category Probability Curves (CPC) (Figure 8.2) to test the assumption of invariance. The peaks of the CPCs appear as a range of hills with distinct peaks and clear crossover points detectable between the curve for one category and the curve for its neighbouring category.

The DELTA raw data thus satisfied the four basic assumptions required for RMM analyses, indicating a significant possibility to be calibrated into linear interval measure.

The dependability of the researchers construct theory is tested with the variable map (see Figure 8.3). Both a top-heavy effect and a ceiling effect were revealed by the variable map. The practical explanation was that the nursing setting where the data were collected facility did not admit acute new patients, but only “stabilised” patients from peripheral hospitals. The ceiling effect can be explained by the state hospital keeping fully functioning patients longer than usual to make sure the patients are proficient with their medication schedules and routine to prevent recurrence. This was confirmed by an additional facility performance report to the facility on length of stays.

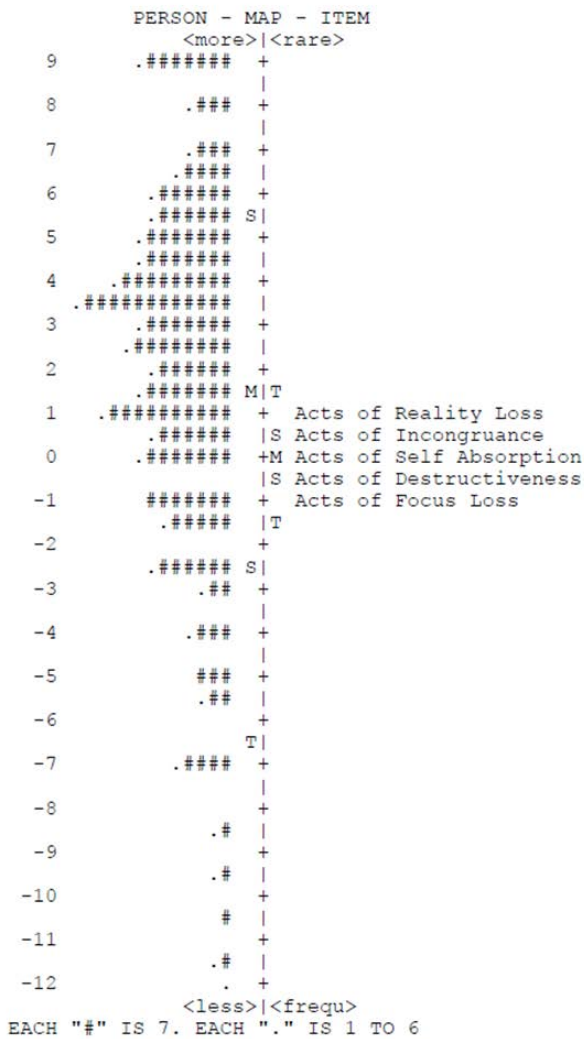


Figure 8.3: Variable map of patient ability and item difficulty in the DELTA.

8.5.4 Results on category functioning

The next step was to select the most basic and most commonly used RMM indices to track the calibration of the categories. Table 8.2 contains the Rasch parameters used to analyse the DELTA category functioning and the results achieved. Linacre (2004) suggested the following eight guidelines to evaluate the category results from the RMM analyses:

- All the DELTA items are orientated with the latent trait variable. **Table 8.3** shows all the items in the Point-Measure Correlation (PT MSE CORR) column with high positive values (>0.89). Linacre warned against negative coefficient values, or values lower than 0.30.
- A minimum of 10 observations is required in each rating category and the DELTA sample fulfils that guideline. However, in the DELTA a low frequency distribution (OBSVD COUNTS) across the first categories (**Table 8.2**) has been observed. This confirms the impact the facility's admission policy has on its patient sample distribution. Patients with very low ability were not frequently observed, as they were first stabilised in another acute facility before transferred to the facility where the data were collected. This was mentioned in the data collection process above.
- The observations are regularly distributed across all rating categories, with higher frequencies in the 3,4,5, and 6th categories. This is also in line with the admission and discharge policies of the facility. Thus, from an operational perspective, the facility where the data was collected have a higher probability of rendering intermediate scores (e.g. 2, 3, 4, 5 or 6) in the dataset, rather than collecting low scores (e.g. 1) on admission or high scores (e.g. 7) on discharge. Importantly, the frequency distribution also correlates with the clinical knowledge where the item that responds earlier on treatment (e.g. focus loss > destructiveness > self-absorption > incongruence > reality loss) reveals a similar pattern. The frequency distribution of the observations follows this clinical logic.

Table 8.2: Results on the DELTA category functioning.

| Item | Category Label | OBSVD COUNT | OBSVD AVRGE | OUTFIT MNSQ | Structure calibration |
|----------------|---------------------------|------------------------|------------------------|------------------------|----------------------------------|
| 1.Reality Loss | 1 | 25 | -10.12 | 0.88 | None |
| | 2 | 143 | -4.94 | 1.05 | -10.82 |
| | 3 | 235 | -.41 | 0.90 | -3.71 |
| | 4 | 266 | 2.14 | 0.87 | -.08 |
| | 5 | 263 | 4.18 | 0.71 | 2.37 |
| | 6 | 153 | 6.05 | 0.68 | 4.63 |
| | 7 | 67 | 7.50 | 1.02 | 7.61 |

| | | | | | |
|--------------------|---|-----|--------|------|--------|
| 2.Incongruous Acts | 1 | 22 | -10.19 | 0.73 | None |
| | 2 | 113 | -5.80 | 0.87 | -10.66 |
| | 3 | 206 | -1.12 | 0.88 | -4.13 |
| | 4 | 291 | 1.75 | 0.83 | -.38 |
| | 5 | 284 | 3.94 | 0.67 | 2.57 |
| | 6 | 159 | 5.98 | 0.63 | 4.98 |
| | 7 | 77 | 7.33 | 0.99 | 7.62 |
| 3.Self-Absorption | 1 | 28 | -10.13 | 0.72 | None |
| | 2 | 77 | -6.48 | 0.67 | -9.43 |
| | 3 | 174 | -1.65 | 1.09 | -4.68 |
| | 4 | 301 | 1.17 | 0.96 | -0.69 |
| | 5 | 290 | 3.55 | 0.83 | 2.68 |
| | 6 | 165 | 5.60 | 0.87 | 5.15 |
| | 7 | 117 | 6.55 | 1.45 | 6.97 |
| 4.Destructiveness | 1 | 25 | -10.04 | 0.88 | None |
| | 2 | 79 | -6.51 | 0.95 | -9.54 |
| | 3 | 174 | -1.88 | 1.02 | -4.51 |
| | 4 | 278 | 1.02 | 0.82 | -0.48 |
| | 5 | 289 | 3.49 | 0.74 | 2.66 |
| | 6 | 166 | 5.37 | 0.71 | 5.17 |
| | 7 | 141 | 6.41 | 1.56 | 6.69 |
| 5.Focus Loss | 1 | 25 | -9.66 | 1.14 | None |
| | 2 | 73 | -6.64 | 1.32 | -8.90 |
| | 3 | 149 | -2.30 | 0.97 | -4.14 |
| | 4 | 219 | 0.42 | 1.06 | -0.28 |
| | 5 | 252 | 2.95 | 1.28 | 2.42 |
| | 6 | 204 | 4.57 | 1.52 | 4.72 |
| | 7 | 230 | 5.48 | 2.16 | 6.17 |

- As required, the average measures advance monotonically with the category as the observed average (OBSVD AVRGE) indicates on **Table 8.2**.
- The OUTFIT MNSQ values for the categories are less than 2.0 as evident on **Table 8.2**. A single value marginally overstepped the recommended guideline, but this was considered insignificant and not warranting remedial work to force these values lower during this level of analysis.
- The thresholds advance orderly with categories as seen in the column Structure Calibration. These thresholds correspond with the intersecting

points between the CPC's in [Figure 8.2](#). It can thus be assumed that the DELTA categories take increasing levels of the latent trait to be observed in higher categories.

- The thresholds are all advancing from neighbouring thresholds by at least 1.0 logits as is evident from the Structure calibration column in [Table 8.2](#). The lowest value being 1.45 logits. This indicates that the categories are clearly separable and functioning independently.
- Only 3 of the DELTA's 30 thresholds are exceeding the 5 logit guideline (5.03, 6.43 and 7.11). These three thresholds all occur between the first and second categories and might be due to the low frequency distribution of category 1 observations in the dataset. It is therefore not considered a structural concern requiring remedial action, unless further analysis provides evidence of a significant influence in the future. The remainder of the thresholds are well between the recommended 1-5 logit ranges, indicating no dead zones in the middle of any category.

In conjunction with the Linacre's ([2004](#)) guidelines above, the CPC's must be evaluated. These curves indicate how the category response structure is predicted to work for any future sample, provided it worked satisfactorily for this sample. Due to a lack of space and to prevent repetition, the CPC of only Item 1: Reality Loss is presented in [Figure 8.2](#). The remaining four items show similar characteristics. The peaks of the seven categories are all in ascending order along the latent variable of each item. At some defined point on the latent variable each category in turn is the most probable than any other one of the categories. Furthermore, the cross-over points between the categories are ordered e.g. the descending curve of each category clearly crosses the ascending curve of the neighbouring category. These cross-over markers are the equal probability points or the thresholds or the parameters of the CPC. This investigation into the categories of the Item 1 ("Reality Loss") concluded that it was ordered and no remedial action for the categories was required. The same finding was evident when investigating the thresholds of the remaining four items.

8.5.5 Results on item functioning

Again, for item analysis, the basic indices most commonly used in the literature were applied. These indices are set up in [Table 8.3](#) and include the infit and outfit mean square (MNSQ) statistics, the point measure correlation (PT MSE CORR), the Rasch reliability on both person and item values and finally the raw variance explain by measures. The raw score analysis of the DELTA item structure ([Table 8.3](#)) showed the similar promising fit results as the raw score category analysis. All the indices are well within the fitting requirements of the RMM.

Table 8.3: Results on the DELTA item functioning.

| Items | Sample | Categories per item | Outfit MNSQ | Outfit MNSQ | PT MSE CORR | Rasch RELIABILITY Person/Item | Variance explained by measure Emp / Mod |
|-------|--------|------------------------|----------------|----------------|----------------|-------------------------------------|--------------------------------------------------|
| 1 | | 7 | 1.38 | 1.38 | .89 | | |
| 2 | | 7 | .97 | .97 | .91 | | |
| 3 | 1152 | 7 | .94 | .92 | .91 | 0.99 / -0.99 | 83.4% / 83.3% |
| 4 | | 7 | .86 | .85 | .91 | | |
| 5 | | 7 | .79 | .79 | .92 | | |

- *INFIT and OUTFIT MNSQ*: This core statistic reports how closely the DELTA corresponds to the Rasch model. With values are around 1, the measure is considered accurate. Thus, with both INFIT & OUTFIT MNSQ values ranging between 0.79 and 1,38 the DELTA's item difficulty range is appropriate to the ability range of the persons under investigation. Consequently, the DELTA can be regarded as a measure with significant levels of accuracy and a high degree of predictability. [Fischer \(2007\)](#) regards these values as reflecting very good to excellent quality in measurement properties.

- The PT MSE CORR reports a noticeably positive correlation of >0.89. This confirms that the distribution and direction from easy to difficult on the DELTA latent variables is aligned with the severity of the patients. Rasch expects the lowest category on the latent variable to be easier for severely ill patients than the highest category.
- *Rasch reliability person/item*: In reliability analysis the RMM quantifies the probability of the DELTA reproducing the same relative location of the measurement point in future applications given the same patients to observe. RMM reports on both person and item reliability, e.g. a "high person reliability" means that there is a high probability that persons estimated with high measurements actually do have higher measurements than persons estimated with low measurements. The same consideration applies to "high item reliability". The DELTA obtained significant person reliability (0.99) and item reliability (0.99) indices. These values were rated by [Fischer \(2007\)](#) as excellent quality in item and person reliability.
- *Variance explained by measure*: This criterion for dimensionality reports empirical and modelled values and must be interpreted as follows: If the data fit the Rasch model perfectly, and the raw variance explained on the empirical values are reported as 83.4%, then that number would have been 83.3%, which is reported as the modelled value. But quality is not only interpreted by how close the empirical and modelled values are, but also how high the percentages are. According to [Fischer \(2007\)](#) values higher than 80% and as close together as the reported values indicate excellent quality in measurement properties.

8.6 Conclusion

The DELTA achieved excellent nursing utility results as a routine nursing measure, as well as very good to excellent instrument quality ratings on the Rasch analysis. Its data satisfied the Rasch fit criteria as a fundamental

measure. No remedial calibration was required to achieve this high rating since it was achieved on the raw data.

CHAPTER NINE

CONCLUSIONS AND RECOMMENDATIONS

Quality in a service is not what you put into it, but what the client gets out of it.

Peter Drucker

9.1 Revisiting the origins and purpose of the study

The subject matter of this study was motivated by the uncompleted work of nursing pioneers. As early as the mid 1800's, Florence Nightingale tried to effect nursing change by focussing on patient outcomes. She pleaded with the authorities of the time to explore beyond the use of mortality statistics; they must also investigate patient outcomes. In 1859 she put a strong case forward for "a uniform system of reporting" on the "proportion of recoveries and average time in hospitals". Without such figures, she wrote, one cannot "argue for change" in nursing care. Despite Nightingale's pleas, 100 years went by before Orlando (1961) reported on "bad" nursing practices as a situation where the patient is not the central character in the nursing process, and where care is directed at nursing goals, not patient outcomes. In 1999, MacVicar pinpointed the cause of the now overdue problem as a "lack of scientific underpinnings or empirical evidence necessary to document the desired outcomes" for patients requiring restorative nursing.

Inspired by these nursing pioneers, this study have attempted to provide the appropriate nursing measures to facilitate change in the way the sub- and non-acute nursing process is conducted in South Africa. If the measures prove to be useful and accurate, they would significantly benefit the emerging restorative nursing sciences to render goal directed services to patients currently having

unpredictable outcomes in sub- and non-acute nursing settings. Secondly, it would provide validated patient-evidence based data to the healthcare management and funding industries in South Africa to do extensive outcomes analysis.

9.2 Conclusions of the study

The study provided significant data and analytic statistics to arrive at scientifically informed conclusions. Below follows the extent to which the results satisfied the expectations and predictions of the study:

9.2.1 Assumptions

The nursing sciences are the glue that keeps all the healthcare services intact. This assumption is based on the nurses' primary caregiver roles, their continuity of presence, and their structured sciences of observing, recording and maintaining patient documentation. Thus nurses have the inherent ability, when given the appropriate tools, to perform the necessary restorative processes in addressing the needs of patients.

However, there are two caveats to this assumption. Nurses will not routinely collect such data if the instrument to collect the data does not add sufficient utility to their nursing process. Furthermore, if the data were to be collected under duress, the data quality would be questionable. Without a reasonable prospect of patient outcomes data collected routinely, the construct validity of any measure will therefore only remain an academic exercise. As a result, the conceptual framework of the study was structured to first determine nursing utility before attempting construct validity. Nevertheless, scientific evidence from the study revealed that nursing utility was confirmed successfully in three of the

four nursing scales. This can be attributed to the collaboration of experienced nurses in the initial development of the scales.

9.2.2 Conceptual framework

The conceptual framework of the study is firmly vested in the theory that for sub- and non-acute patients, the extent of their functional gain is dependent on the restorative nursing performance they received. This concept formed the central theme that connected all the aspects of the inquiry coherently throughout this study. Therefore, if the patient's functional status can be longitudinally quantified over the nursing days, the patient improvement between admission and discharge can be numerically calculated. With empirical evidence of patient outcomes available, one can infer effective and efficient nursing performance, leading to the assumption that the quality of restorative nursing can be directly related to patient outcomes.

In [Annexure J](#) the conclusion to the study's conceptual framework is illustrated with the basic calculations done on the DELTA data collected at the pilot facility. This method of performance reporting can be made available to the facility as required as the calculations can be electronically generated on software where the data is captured. Firstly, nine mental healthcare diagnostic groupings were identified in the facility's dataset. When the DELTA scores were analysed in these diagnostic groupings it became clear that each group had a uniquely different outcome pattern when comparing their admission scores, discharge scores and length of stays (LOS). These three core indicators are the stepping stones to calculate the effectiveness and efficiency performance benchmarks. Effectiveness refers to the functional gains (discharge score minus the admission score) and efficiency refers to the functional gains divided by the average length of stay (ALOS = days) in the facility. The basic calculations in [Annexure J](#) showed that different impairment groups have different outcomes. However, if postulated differently: *the same patients within the same impairment group will have the same outcomes given the same treatment*, it provides the

theoretical framework for future peer review analyses of patient outcomes and nursing performance.

Most importantly for this study, it is concluded that patients with previous unpredictable outcomes now have predictability. From [Annexure J](#) one can now predict that a patient with e.g. a mood disorder, and with a total DELTA score of 17 / 35 on admission, will be discharged after 25 days of restorative nursing care with a DELTA score of 28,4 / 35. The patient outcome is therefore effectively 11.4 DELTA scores and the efficiency ratio of the nursing is 0,61 (11.4 divided by 25), meaning the nurses are performing at a rate of 0,61 DELTA scores gains per day in patients with mood disorders.

The same principle and calculations would apply to the BETA and GAMMA analysis. Only different impairment groups for each scale would be used. Nursing performance ratios would therefore differ between impairment groups and different scales. The restorative nursing performance ratios for strokes might be 0,31 BETA scores gain per patient day and for post hip replacement 2,4 BETA scores gain per patient day. Furthermore, with closer analyses of the nursing performance ratios and patient outcomes one might find that some nursing teams are better skilled in specific impairment groups e.g. orthopaedic rather than neurological cases.

As more facilities would use the measures, the facilities data can be pooled, the performance analyses per measure (e.g. DELTA) can provide considerable comparative statistics in a peer review format. The national averages can become the outcomes' benchmarks. Furthermore, as nurses apply their restorative techniques into nursing sciences and continuously increase their performance (e.g. DELTA gains per day) so would their patients' outcome advance and their predictability would improve.

9.2.3 Research objectives

The study provided validated evidence that the two main objectives of the study were successfully achieved. Significant levels of nursing utility were evident from the descriptive data analysis concerning the use of the BETA, GAMMA and the DELTA. The ALPHA however was rejected by practicing nurses as not having nursing utility, and thereby the ALPHA was excluded from further construct validity analyses. However, as the three remaining measures achieved acceptable levels of nursing utility, they were subjected for further analyses to the RMM. They subsequently satisfied the RMM fit statistics and achieved ratings between good and excellent, which indicate their potential to be calibrated into standardised fundamental nursing measures.

9.2.4 Research problem

The research problem was to search for an appropriate scientific method to provide evidence of validity of the proposed nursing scales. A valuable secondary consideration would be to also provide empirical evidence of the scales' potential to be transformed from ordinal scales to linear interval measures. A suitable research method was found as reported in Chapter Four. This scientific method followed a similar deductive approach for all the scales as reported in Chapters Five, Six, Seven and Eight. It consisted of a chain of interdependent methods to finally satisfy the two main objectives, namely nursing utility and construct validity. The assumed validations of the four scales were scientifically tested and the ALPHA was discarded as not adhering to either nursing utility or fundamental measurement standards. The remaining three measures satisfied these criteria in different rankings.

9.2.5 Research purpose

The research intention was that the nursing measures would change the way sub- and non-acute nurses go about caring for their patients who require restorative care. On the question whether these measures have the potential to make nurses responsive to patient needs for restorative nursing care, the answer is an unambiguous yes. Significant evidence was provided that nurses accepted the measures as useful and are implementing it within their nursing care plan routinely with very good results. The evidence also concluded that nurses are actively attempting restorative techniques to improve the function of their patients. The RMM also provided fit statistics that the raw data can be calibrated into validated measurements.

9.3 Reflections

Over and above the discussions rendered frequently throughout the study, the following implicit reflections are recorded explicitly as they were noteworthy to the researcher.

9.3.1 Nursing intuition

Most important to the successful outcome of the study was the inclusion of experienced practising nurses early in the development of the measures. Their knowledge of nursing practice not only created a hypothetical scenario of clinical utility, but their insights into the observable stepping stones of a patient's functional gains during a recovery period, facilitated the process of rating scale development. Therefore, by exploring nursing intuition upfront, and applying its clinical richness, a platform of knowledge for the design and development was created. This explorative experience provided the researcher with enormous

insights into the hidden opportunities the primary caregiver's scope of practise has to offer to the healthcare industry as a whole, as can be seen in the DELTA outcomes in [Annexure J](#).

The nurses' closeness to the basic elements that matter most in realising a better outcome for their sub- and non-acute patients and being oblivious of their potential to influence this outcome was a privilege to unlock and redefine. Their sudden awareness of the valuable role that the primary caregiver can play to improve patient outcome, became clear when daily empirical evidence became available to show how little resourcefulness from the nursing side can significantly improve the patient's functional scores. This new reality to patient care became a personal challenge to their inventiveness as nurses to improve their patients' independence over time, and it dawned on them that functional independence is what their nursing outcomes was all about.

The conflicting components to this discussion is that the measures were based on nursing intuition, but only when implemented did they find it an awakening experience towards patient and nursing outcomes. The explanation lies in the nurses' underestimation of their nursing intuition as not being a dependable resource. The evidence to their underestimation became apparent when they demonstrated a measure of surprise that the Rasch analysis of the measures confirmed the accuracy of their intuition or their "sixth sense". It was clear that not only did they misapprehend the accuracy of their contribution to the measurement development, but also their significant contribution to patient outcomes.

9.3.2 Sustainability

A driving force behind the study was to achieve sustainability. Although nursing utility has been the study's parameter to infer sustainability, there is more to sustainability than the degree of conviction the nurses have about the

usefulness of the instrument in their practice. The drive for sustainability cannot only come from the nursing community, but should also come from the spin-offs and the secondary beneficiaries of having a validated nursing measure providing routine patient outcome data.

Beneficiaries to patient–evidence based outcomes data are the healthcare risk management and funding industries. Valuable calculations as set out in [Annexure J](#) are required by healthcare funders to establish some predictability in an environment where patients were labelled as being unpredictable. When these industries would become aware of the availability of the accurate nursing data to calculate predictability, the patient files might become a sought after commodity that might further sustain the on-going use of the nursing measures.

Another beneficiary that could add sustenance to the longevity of the measures is the nursing auditing process. If it can be concluded that Table 9.1 infers equally to patient outcomes and nursing performance, then the effectiveness and efficiency of nursing service delivery can be empirically quantified, allowing the quality of nursing also to be managed on outcomes parameters and not only on input parameters. Outcome-based nursing audits would significantly enhance the sustainable use of the nursing measures.

9.3.3 Restorative nursing domain

There is factual evidence that nurses, when becoming aware of their inherent abilities to influence patient function towards independence, and being able to quantify their ingenuity, are now experimenting to include restorative nursing concepts into the nursing process. Moreover, this change is currently happening without any formal or scientific guidance. It is happening spontaneously purely because the measurements made the nurses aware of their patients' potential to regain independence and the nurses have discovered their own potential to facilitate this process as restorative nurses. It can be concluded that the nurses' understanding of the traditional needs assessment of a patient has changed.

It does not only refer anymore to the nursing tasks to render comfort and healing, but to a greater extent it now also includes the nursing techniques to render active restoration towards independence. This instinctive evolvement from basic nursing care towards restorative nursing, whilst tracking patient outcomes from day to day, is an encouraging phenomenon. The next natural steps would be to document the current nursing techniques and practices as the primary steps to the development of a scientific based restorative nursing science for sub- and non-acute nursing practices in South Africa.

9.3.4 Paucity of knowledge

Neither in the literature review nor in personal communication with prominent South African nursing academics could any evidence be found of validated nursing measures used for routinely observing human function. Unless any omission is found later, these nursing measures would be the first documented nursing measures to fill this gap in executing the nursing process. The uniqueness to the BETA, GAMMA and DELTA, over and above the rarity of similar measures in the nursing sciences, is that they measure patient functionality and therefor renders patient-evidence based outcomes data.

Patient-evidence based outcomes data, as opposed to practice-evidence based data, opens wide a range of research opportunities to the nursing sciences. For the first time sub-and non-acute nursing care can apply validated empirical variables to model patient outcomes. The table [Annexure J](#) is an example how pooled patient-evidence based data can be manipulated with basic statistics and create powerful dependent variable norms and benchmarks to understand patient outcomes. Quantifiable goals can be set on admission and trends in nursing performance can easily be monitored. Furthermore, to add value to future restorative nursing research, the impact of newly founded nursing techniques (as independent variables) can be tested against the modelled patient outcomes (dependent variable).

9.3.5 Routine data

Throughout the study it was found that nurses were applying the nursing measures routinely and used the scores in a uniform language to describe how patients improve or decline. In the case of the BETA they also recorded the scores daily in the patient files, in the DELTA the scores were recorded weekly, and in the case of the GAMMA it was recorded as changes were observed. The nurses thus monitored patient outcomes routinely and recorded the scores in the patient files. The patient files are now rich with empirical longitudinal patient level data which provide significant new insights into the sub-and non-acute healthcare sciences.

Internationally, [Purkis *et al* \(2005\)](#) reported that there is an increasing emphasis to provide routine outcomes measurement at patient level for sub-and non-acute nursing care settings. A healthcare outcome is a measurable change in a patient's health as a result of a healthcare intervention. In its purest form, a measurable healthcare outcome implies the measurement of a patient's health status before an intervention is carried out, and measuring the health status again after the intervention has been completed and then compare the measurements with the intervention rendered. It follows that to successfully achieve healthcare outcome the measures must be done and recorded routinely. This study revealed that the measurements are done routinely and recorded in the nursing documentation and the measures are validated. This infers that outcomes measurements in sub-and non-acute care settings can now be achieved successfully.

Valuable longitudinal data based on patient improvement patterns became available to the data analysts. In the hands of healthcare statisticians, this evidence of nursing effectiveness is required to plan, implement and monitor the new reforms in healthcare services. This high level manipulation of data to generate policy documentation is the same data that the humble primary caregivers collected whilst observing their patients functioning. It is also the

same basic data that drives the caregiver to find new techniques to improve their patients' functional score.

9.3.6 Patient outcomes

The nurses' acceptance to implement, observe and record patient function routinely into their nursing care plans also inferred buy-in from the nurses to accept ownership of the patient level outcome scores. Accepting ownership was a milestone in the turn-around process from task driven nursing towards patient outcomes driven nursing. From this point it was evident that as a natural consequence, they accepted their role as restorative nurses and are prepared to routinely monitor their patients' outcomes against their restorative inputs. There is evidence that, to the nurses, the collection of patient-evidence based data now became a secondary consequence; their primary concern was exploring and discovering new restorative techniques to gain patient functionality and to provide empirical proof of their new achievements.

The primary beneficiary of this new process is the patient with previously unpredictable outcomes. The new focus is now on the patient and their functional status and nursing effectiveness is now measured by their improvement over time. The faster the gains, the better restorative nursing. Literature reviews mention early restorative interventions to be the most important independent variable to successful patient outcomes. Although restorative nursing has a long way to go in South Africa, the early signs of accepting ownership and accountability are very promising for patient outcomes.

9.4 Contribution to nursing knowledge

The study made a significant contribution to the sustainable gathering of patient

level outcome data which were previously not available to the South African nursing and healthcare sciences. Nurses at large were not aware that they have this ability to accurately provide empirical data on patient functioning. This study unlocked this restricted view and allowed nurses to move to higher levels of professional fulfilment. Being able to measure patient improvement and being able to place that measurement ability within the reach of the level of nurses where the most difference can be accomplished, is a substantial contribution. By making it part of their initial training the new generation nursing assistants would be capable to provide restorative care nursing services and provide data of their progress. Professional nurses in managerial positions will have to find new methods of managing nursing assistants without stifling their enthusiasm, but also encouraging the new found eagerness in managing their patient functional scores upwards. It is anticipated to be a daunting situation for the traditional task driven nursing professional, but the evidence from the study concluded that it is a “wonderful” experience for the nursing professional interested in patient outcomes.

To date, numerous applications for the use of the nursing measures in facilities have been received. Corporate actuarial companies are also seeking assistance on how to apply the nursing data in constructing outcomes models for unpredictable patients. The opportunity of actuarial analyses to develop clinical governance models based on patient-evidence data is unlimited, and it has significant healthcare funding and policy implications. However, it is not within the scope of this study to postulate on the outcome of the study might have on funding and policy ventures. Suffice it to say one must never forget or trivialise the conscientious role played by the modest professional nurse, nursing assistant or caregiver that observes and scores and provide the data required for these far reaching implications. It must also be remembered that it is they who will provide both the change towards independence of the patient with an unpredictable outcome and the evidence of doing so.

9.5 Recommendations for nursing practice

To implement restorative nursing in sub- and non-acute care settings in South Africa, minor changes to the current nursing care plans and processes would result in major positive patient outcomes. These recommended changes in nursing practice can be listed as follows:

- In their basic training all nurses must be made aware that the purpose of sub- and non-acute nursing care is to achieve functional independence for the patients. As patient functional change is now measurable, the outcomes of patients in sub- and non-acute care can be expressed empirically. Through the educational processes nurses must also be made aware of the relationships between nursing inputs and patient outcomes, and that sub- and non-acute nursing inputs must be restorative in nature to achieve patient outcomes. The restorative nursing techniques must be directed to improve the patient's outcomes as inferred by the scores. Therefore nurses must be taught to apply the nursing measures and the most effective techniques to improve outcome scores in the quickest and most sustainable manner. If this concept is embedded into the nursing process the beginnings of restorative nursing would become evident.
- Nursing assistants and caregivers, who currently render all the patients' ADLs, must be recognised and trained as the primary providers of ADL restorative nursing techniques. Their scope of practice must be adapted to include restorative activities. The study has identified that these level of nurses, because of their proximity to the ADLs, experienced the change in the nursing approach from task rendering to restorative nursing as a natural occurrence.
- The nursing process must be adjusted to observe, score and record the patient-evidence based scores routinely. Ideally the patient scores must be part of the patient care plan and recorded in the patient file by the nurse that does the scores.

- Once the above has been agreed on, the task of the professional nurse would be to:
 - Establish empirically the functional status of the patient on admission. This set of scores would serve as a baseline assessment from which the restorative nursing process is designed.
 - Predict an empirical outcome for the patient by setting goals to be achieved. This can be done either through experience, or using the predicted benchmarks rendered by data analysis. This prediction is taking into consideration “similar patients will have similar outcomes given similar restorative nursing”. With the predicted goals set, as shown in **Figure 9.1**, the restorative nursing process can be implemented and the nursing performance (e.g. score change per day) can be monitored.

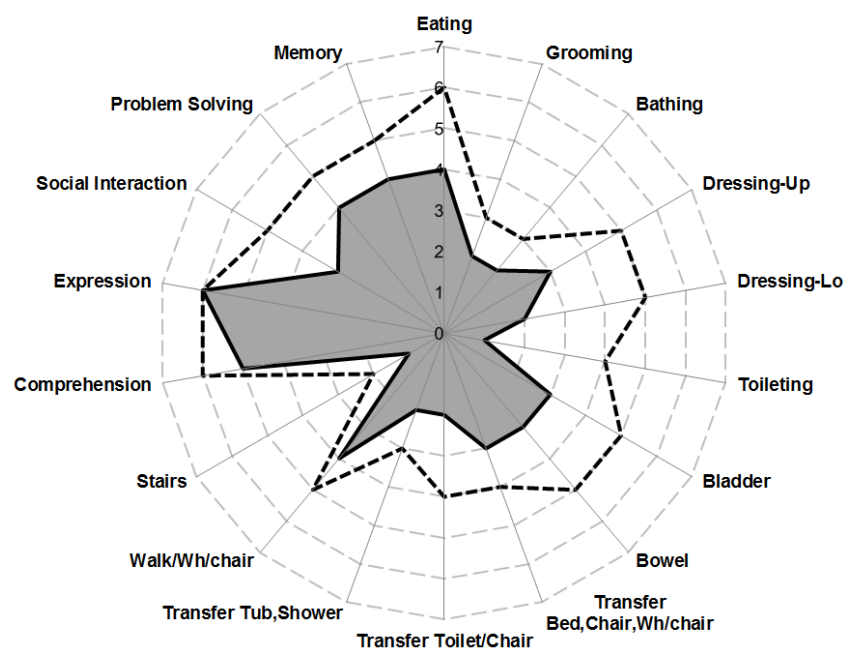


Figure 9.1: BETA scale with admission and predicted outcome score

- If the nursing performance, as empirically monitored by the score change per patient day ratio, is exceeding the predicted

improvement rate for the diagnostic group the patient belongs to, (e.g. stroke) the professional nurse must investigate if the nursing assistants are scoring correctly. If the scoring was correct, then the professional supervising nurse must investigate and record the restorative techniques applied to achieve the higher performance. If the score change per day is below predicted improvement rate, a similar procedure is followed. If found that the restorative nursing technique does not render successful results, the professional nurse would consult therapists for assistance to overcome the barrier to nursing performance. These very basic models will firmly embed the restorative sciences in the nursing process.

- Finally, included in the professional nurse's supervision of the nursing assistants and caregivers is a weekly reflective counselling sessions to debrief them on problem situations encountered, new restorative nursing techniques found and other possible solutions to improve patient outcomes. The objective must always be to improve the nursing performance which is reflected in the score change per patient day ratio.
- To support the restorative nursing process, software must be made available to the nurses where patient scores could routinely be entered into an electronic database. Once in the system, various automated calculations become available to the restorative nurse.
 - The patient diagnosis and admission score can immediately provide the nurses within the predicted outcome of the patient in terms of the anticipated discharge score and average length of stay. These predictions will be based on the facility's track record of restorative nursing performance of similar cases.
 - The data from the facility can be electronically offloaded to a national platform and actuarial analysis would provide nurses to evaluate their performance against national benchmarks of patient outcomes and nursing performance.

- Over and above the advantage of restorative nurses competing with themselves on nursing performance, the performance ratios (scores gain per patient day) can also benefit the nursing audit process. Nursing teams with high performance ratios clearly have skills that nursing teams with low performance ratios need. The auditing process should ideally incorporate a skills transfer process from those that have the skills to those that require the skills.

9.6 Recommendations for future research

The methodology used in this study was proven successful and would set the agenda for various research models for into restorative nursing outcomes. Solid platforms for on-going research to improve patient outcomes and nursing performance created. The first objective for future research would be to analyse the data and find patient groupings with similar outcomes on which to base the predictability models on. Thereafter these patient groupings must be investigated to understand the independent variables in the groupings that still create unpredictability e.g. the outliers such as co-morbidities (arthritis, heart failure, emphysema etc.) and co-disabilities (previous stroke, head injury etc.).

A situation that has been mentioned before but not been concluded in this chapter is the question that often confronts a researcher when investigating clinical utility and psychometric properties jointly: What should be do when a scale's nursing utility proves to be excellent in clinical decision making, but the statistical analyses of construct validity requires adjustments? Should the scale's rating algorithms be redesigned to conform to the statistical findings and thereby risk the high levels of nursing utility already achieved?

To prevent a possible revisit to nursing utility and confusion, it is suggested to retain the existing algorithms. It is best to leave it to modern technology to find a solution to rectify the problem.

This can be done when the raw scores are electronically captured weekly in a web-based software application, as it is already done. The RMM recommended changes can then be done electronically in the software providing adjusted RMM data only in percentage format to the nurses, thus not adding to confusion. Thus, at caregiver level the nurses can continue to use the raw scores as it makes clinical sense and provide a framework for implementing restorative nursing techniques, but once the raw scores are entered into the software, the accurate patient outcomes will be available to them as RMM percentages.

9.7 Final note

This study was an attempt to create a better space for sub-acute and non-acute patients and their nurses. There is enormous satisfaction that the research was successful in achieving the intent. However, on a personal level, the learning curve into the intricacies of the scientific research and finding the right way to put the puzzle of variables together was an exhilarating and rewarding experience. Now, at the end of this long road, I find myself also in a better space.

One of the greatest discoveries a man makes, one of his greatest surprises, is to find he can do what he thought he could not do.

Henry Ford.

Bibliography

AANP. (2011). *The American Academy of Nurse Practitioners*. Accessed 30 December 2011. [www.luc.edu/nursing/pdfs/APN_certification2011.pdf]

Aas IHM. (2010). Global Assessment of Functioning (GAF): properties and frontier of current knowledge. *Annals of General Psychiatry*. 9:20. doi: 10.1186/1744-859X-9-20. [[PMC free article](#)] [[PubMed](#)] [[Cross Ref](#)]

Aas IHM. (2011). Guidelines for rating Global Assessment of Functioning (GAF) *Annals of General Psychiatry*. 2011;10:2. doi: doi: 10.1186/1744-859X-10-2. [[PMC free article](#)] [[PubMed](#)] [[Cross Ref](#)]

Acello B. (2009). *The Long-Term Care Nursing Desk Reference*. HCPco, Inc, Marblehead. MA. USA.

ICF (2001) International classification of functioning, disability and health: ICF short version. World Health Organisation Library cataloguing-in-publication data.[ISBN 92 4 154544 5].

Adams M, Palmer A, O'Brien JT & Crook W. (2000). Health of the nation outcome scales for psychiatry: Are they valid? *Journal of Mental Health*. Vol 9 p 193–198.

American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurements in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Anderson Dr S. (2010). Personal Interview: Existence of nursing measures on human functionality. 15 January 2010. Gauteng Dept of Health.

Anderson RI, Buckwalter KC, Buchanan RJ, Maas MJ & Imhof SI. (2003). Validity and reliability of the minimal data set depression rating scale (MDSMRS) for older adults in nursing homes. *Age and Aging*. Vol 32 p 435-438.

Andrich D. (1978). A rating formulation for ordered response categories. *Psycho-metrika*. Vol 43 p 357-374.

APA. American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders, fourth edition*. Washington, DC: American Psychiatric Association.

APA Task Force on Evidence Based Practice. (2006). Report of the 2005 Presidential Task Force on Evidence-Based. *American Psychologist*. Vol 61 p 271-285.

Bandeem- Roche K, Xue QL, Ferrucci L, Watson J, Guralnik JM, Chalves P, *et al.* (2006). Phenotype of Frailty: characterisation of in women's health and aging studies. *Journal of Gerontology: Series A Biological Sciences and Medical Sciences*. Vol 61:3 p 262-266.

Barbara A & Whiteford G. (2005). Clinical utility of the handicap Assessment and Resource Tool: an investigation of its use with people in hospital. *Australian Occupational Therapy Journal*. Vol 52:1 p 17-25.

Bebbington P, Brugha T, Marsden L & Window S. (1999). Validation of the Health of the Nation Outcomes Scale. *The British Journal of Psychiatry*. Vol 174 p 389-394.

Bialocerkowski AE & Bragge P. (2008). Measurement error and reliability testing: Application to rehabilitation. *International Journal of Therapy and Rehabilitation*. Vol 15:10 p 422.

Billay D, Myrick F, Luhanga F & Yonge O. (2007). A Pragmatic View of Intuitive Knowledge in Nursing Practice. *Nursing Forum*. Vol 42:3 p 147–155.

Bingham R. (1943). "The Kenny Treatment for Infantile Paralysis: A Comparison of Results with Those of Older Methods of Treatment." *The Journal of Bone and Joint Surgery* 25: p 647–650. Online: [www.ejbs.org/cgi/reprint/25/3/647.pdf.]

Black J, Lewis T, McIntosh P & Callaly T. (2009). "It' not that bad: the views of the consumers and carers about routine outcome measurement in mental health. *Australian Health Review*. Vol 33:1 p 93-99.

- Blaum CS, O'Neill EF, Clements KM, Fries BE & Fiatarone MA. (1997). Validity of the minimal data set for assessing nutritional status in nursing home residents. *American Journal of Clinical Nutrition*. Vol 66 p 787-794.
- Bond TG & Fox CM. (2007). Applying the Rasch Model: Fundamental Measures in the Human Sciences. *Mahwah NJ. Lawrence Erlbaum Associates*.
- Brink H. (2008). *Fundamentals of research methodology for healthcare professionals*. Mills Litho, Maitland, Cape Town South Africa.
- Brink W. (2010). Food/Cooking/Nutrition/Community. (Accessed 15 June 2011) [<http://www.thefreelibrary.com/Sarcopenia+Causes,+Prevention,+and+Treatment.+Part+1-a01074034085>]
- British Department of Health. (1998). *Modernising Social Services*. London. TSO.
- Brogden HE. (1977). The Rasch Model, the law of comparative judgement and additive conjoint measurement. *Psychometrika*. Vol 42 p 631-634.
- Brunero S & Stein-Palbury J. (2008). The effectiveness of clinical supervision in nursing: An evidence based literature review. *Australian Journal of Advanced Nursing*. Vol 25:3 p 36-44.
- Campbell NR. (1919). *Physics: The elements*. Cambridge University Press.
- Campbell NR. (1921). *What is science?* Methuen and Company Ltd. London.
- Carey M. (1994). The group effect in focus groups: planning, implementing and interpreting focus group research. In: *In Critical Issues in Qualitative Research Methods (Morse J., ed)*, Sage Publishers, London. p 225-241.
- Certificate of Active Participation. Issued By Dr John Linacre, Editor, *Rasch Measurement Transactions*. 27 September 2010. Copy available from researcher.
- Clarke Prof M. (2010). Telephonic Interview: Existence of nursing measures on human functionality. 17 January 2010. University of Stellenbosch: Dept of Nursing.

CMS (Centres for Medicare & Medicaid Services). (2010). (Accessed 16 December 2010).

[www.cms.gov/NursingHomeQualityInits/45_NHQIMDS30TrainingMaterials].

Cochrane A. (1972). Effectiveness and Efficiency. Random Reflections on Health Services. *London, Nuffield Provincial Hospitals Trust.*

Cook ET. (1913). *The Life of Florence Nightingale*. London: Macmillan.

Crookes VA & Andrews GL. (2009). Primary healthcare: people, practice and places. *Ashgate Publishers, Surrey, England.*

Dale AE. (2005). Evidence-based practice: compatibility with nursing. *Nursing Standard*. Vol 19:40 p 48-53.

Daniels R. (2004). Nursing Fundamentals: Caring and Decision making. *Delmar Learning, Clifton Park New York 10265*. [www.delmarlearning.com].

David FN. (1962). *Games, Gods and Gambling. A History of Probability and Statistical Ideas*, London: Charles Griffin & Co. Ltd. Reprinted by Dover Publications, 1998.

Dilorio C, Hockenberry-Eaton M, Maibach E & Rivero T. (1994). Focus groups: an interview method for nursing research. *Journal of Neuroscience Nursing*. Vol 26:3 p 175–180.

Eager K & Innes K. (1992). *Creating a Common Language: The Production and Use of Patient Data in Australia*. Commonwealth Department of Health, Housing, Local Government and Community Services. AGPS.

Eagar K, Gordon R, Hodkinson A, Green J, Eagar L, Erven J, Eckstein G, Spooner G, Kennedy C, Owen A, Cromwell D, Leotta T & Rigley A. (2001). *The Australian National Sub-Acute and Non-Acute Patient Classification (AN-SNAP): report of the National Sub-Acute and Non-Acute Case-mix Classification Study*. Centre for Health Service Development, University of Wollongong, Sydney, Austraria

Eisen SV & Dickey V. (1996). Mental Health Outcomes Assessment: The new agenda. *Psychotherapy*. 33 p131-189. Gobet F & Chassey P. 2008 Towards an

alternative to Brenner's theory of expert intuition in nursing: a discussion paper. *International Journal of Nursing Studies*. Vol 45:1 p129-139.

Elo S & Kyngäs H. (2007). The qualitative content analysis process. *Journal of Advanced Nursing*. Vol 62 p 107–115.

Embretson SE & Reise SP. (2000). *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum.

Fawcett A. (2007). *Principals of assessment and outcome measurement for occupational therapists and physiotherapists. Theory, skills and application*. John Wiley & Sons Ltd, West Sussex. England.

Fisher SE, Bergio ID, Thorn BE, Allen-Brugge R, Gerstle J & Roth DL. (2002). Pain assessment and management in cognitively impaired nursing home residents: Association of certified nursing assistant pain report, minimal data set pain report, and analgesic medication use. *American Geriatric Society*. Vol 50 p 152-156.

Fischer WP. (1993). Measurement-related problems in functional assessment. *American Journal of Occupational Therapy*. Vol 47 p 331-338.

Fisher WP Jr. (2007). Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions*. Vol 21:1 p 1095.

Fischer WP Jnr. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*. Vol 21:1 p 1095.

Flintoft VF, Williams JI, Williams RC, Basinski AS, Blackstien-Hirsch P & Naylor CD. (1998). "The need for acute, subacute and nonacute care at 105 general hospital sites in Ontario. Joint Policy and Planning Committee Non-Acute Hospitalization Project Working Group." *Canadian Medical Association Journal*. Vol 158:10 p 1289-1296.

Foxcroft CD, Patterson H, Le Roux N, & Herbst D. (2004). Psychological assessment in South Africa: a needs analysis: the test use patterns and needs of psychological assessment practitioners: final report. July. *Assessment*

technology and evaluation. Human Sciences Research Council Web site: [http://www.hsrc.ac.za/research/npa]. Accessed September 2010.

Gibson L & Strong A. (1997). A review of functional capacity evaluation practice. *Work*. Vol 9 p 3-11.

Gibson L & Strong A. (2002). Expert review of an approach to functional capacity evaluation. *Work*. Vol 9 p 231-242.

Gillies DA. (1982). *Nursing management: a systems approach*. Philadelphia, PA: WB Saunders Co.

Graf C. (2007). The Lawton Activities of Daily Living (IADL's) Scale. *Best Practices to Nursing Care to Older Adults*. Vol 23.

Granger C. (2011). Quality outcome measures for rehabilitation programs. Medscape. Accessed 10 October 2011. [emedicine.medscape.com/article/317865-overview]

Heinemann AW, Kirk P, Hastie BA, Semik P, Hamilton BB & Linacre JM. (1997). Relationships between disability measures and nursing effort during rehabilitation for patients with traumatic brain and spinal cord injury. *Archives of Physical Medicine and Rehabilitation*. Vol 78:2 p143–149.

Hellerstein DJ. (2008). Practice-based Evidence rather than Evidence-based Practice in Psychiatry, *Medscape Journal of Medicine*. Vol 10:6 p 141 [http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2491680/]

Hendrix C C, Sakauye K M, Karabatsos G & Daigle D. (2003). The use of minimal data set to identify depression in the elderly. *Journal of the American Medical Directors Association*. Vol 4 p 308-312.

Herbst Prof MC. (2010). Personal Interview: Existence of nursing measures on human functionality. 10 January 2010. Unisa: Dept of Health Sciences (Emeritus).

Heyns Dr T. (2010). Telephonic Interview: Existence of nursing measures on human functionality. 14 January 2010. University of Pretoria: Dept of Nursing.

Hsieh H & Shannon S. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research Journal*. Vol 15:9 p 1277 -1288.

Human Prof S. (2010). Telephonic Interview: Existence of nursing measures on human functionality. 17 January 2010. Unisa: Dept of Health Sciences.

ICF Accessed 10 June 2011 [<http://www.who.int/classifications/icf/en/>].

Innes E & Straker L. (2003). Attributes of excellence in work related assessments. *Work*. Vol 20 p 63-76.

Innes E & Striker L. (2003a). Attributes of excellence in work related assessments. *Work*. Vol 20 p 63-76.

Innes E & Striker L. (2003b). Workplace assessments and functional capacity evaluations: Current beliefs of therapists in Australia. *Work*. Vol 20. p 225-236.

Iramaneerat C, Smith EV & Smith RM. (2008). Chapter four in: Osborne JW. *Best practices in quantitative methods*. SAGE Publishing; Thousand Oaks, California.

IRF – PAI. (2001). *Inpatient Rehabilitation Facility Assessment Instrument Training Manual. Uniform Data System for Medical Rehabilitation*; Buffalo, New York. Accessed 20 August 2010. [www.hcfa.gov/medicare/irfpps.htm]

James C, MacKenzie C & Higgenbotham N. (2007). Health Professionals' Attitudes and Practices in relation to Functional Capacity Evaluations. *Work*. Vol 29:2 p 81-88.

James C & MacKenzie L. (2009). The clinical utility of functional capacity evaluations: The opinion of health professionals working within occupational rehabilitation. *Work*. Vol 33 p 231-239.

Joffe B. (2009). *My Book*. 3ed. Phoenix design. Fontainbleau . South Africa

- Kitzinger J. (1996). Introducing Focus Groups. In: *Qualitative Research in Health Care (Mays N. & Pope C., eds)*, B. M. J. Publishing Group, London. p 36–45.
- Kopf EW. (1916), "Florence Nightingale as statistician," *Journal of the American Statistical Association*, Vol 15 p 388 - 404.
- Kottorp A. (2003). Occupational based evaluation and intervention. Validity of the assessment of motor process skill when used with persons with mental retardation. *Umea University, Sweden*.
- Krueger RA. (1994). Focus Groups, *A Practical Guide for Applied Research*, 2nd edn. Sage Publications, Thousand Oaks, p 36–45.
- Krueger RA & Casey MA. (2000). *Focus groups: A Practical Guide for Applied Research* (3rd edition). Thousand Oaks. California : SAGE.
- Kunik CL, Flowers L, Kazanjian T. (2006). Time to rehabilitation admission and associated outcomes for patients with traumatic brain injury. *Archives of Physical Medicine*. Vol 87:12 p 1590–1596.
- Lakeman R. (2004). Standardised routine outcome measurement: Pot holes in the road to recovery. *International Journal of Mental Health Nursing*. Vol 13 p 210-215.
- Lang TA, Hodge M, Olsen V, Romano PS, Kravits R. (2004). A systematic review of the effects of nurse staffing on patient, nurse and hospital outcomes. *JONA's Healthcare Law, Ethics and Regulation Journal*. Vol 7:8 p 326-337.
- Lawton MP & Brody EM. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist*. Vol 9 p 179-186.
- Le May A & Williams A. (2006). Indicators of effectiveness – exposing the range, robustness and relevance of outcome measures for the evaluation of nursing and midwifery practice: A literature review. *Clinical Effectiveness in Nursing*. Vol 951 p 113–131.

Likert R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*. Vol 140 p 1-55.

Linacre JM & Wright BD. (1989). The “length” of a logit. *Rasch Measurement Transactions*. Vol 3:2 p 54-55.

Linacre JM. (2002). Optimising Rating Scale Category Effectiveness. *Journal of Applied Measurement*. Vol 3:1 p 85-106.

Linacre M. (2004). Optimising rating scale category effectiveness. In E V Smith Jr and RM Smith (Eds). *Introduction to Rasch measurement: theory, models, and applications* (p 258 – 275). Maple Grove, MN: JAM press.

Linacre JM. (2005). Measures, Meaning and Morality. Pacific Rim Objective Measurement Symposium (PROMS) & International Symposium on Measurement and Evaluation (ISME). [www.rash.org/memo71.pdf] Accessed 20 November 2010.

Linacre M. (2010). Practical Rasch Measurement – Topics Course. http://www.statistics.com/how_coursework. Retrieved from internet from 20 August to 30 September 2010. Copies in custody of researcher.

Linacre M. (2010) #2. Practical Rasch Measurement – Topics Course. Lesson 1. http://www.statistics.com/how_coursework. Retrieved from internet from 20 August to 30 September 2010. Copies in custody of student.

Linacre M. (2010) #3. Practical Rasch Measurement – Topics Course. Lesson 3. http://www.statistics.com/how_coursework. Retrieved from internet from 20 August to 30 September 2010. Copies in custody of student.

Linacre JM. (2011). A user’s guide to WINSTEPS (3.70.1.1 ed.) *Chicago: MESA. Press*.

Loubser HJ & Raath C. (2004). SADFM levels of care analysis. *Unpublished study done for the South African National Department of Health*. Copy in custody of student.

Luce RD & Tukey JW. (1964). Simultaneous conjoint measurement. *Journal of Mathematical Psychology*. Vol 1 p 1-27.

Lum TY, Lin WC & Kane KL. (2003). Use of proxy respondents and accuracy of minimal data set assessments of activities of daily living. *Journal of Gerontology. Series A. Biological Sciences and Medical Sciences*. Vol 60 p 654-659.

Lundgren-Nilsson A. (2006). *Characteristics of the Functional Independence Measure in patients with neurological disorders*. PhD Thesis. Gotenborg University. Sweden.

Lundgren-Nilsson AL.(2006). *Characteristics of the Functional Independence Measure in patients with neurological disorders. Measurement through Rasch analysis*. PhD Study. Institute of Neurosciences and Physio/Rehabilitation Medicine, Sahlgrenska Academy, Gotenborg. Sweden.

Lynch G S. (2004). Tackling Australia's future health problems: developing strategies to combat sarcopenia – age-related muscle wasting and weakness. *Internal Medicine Journal*. Vol 34 p 294–296.

Mantle J & Funke-Furber J. (2003). (In dialogue with Verah McIver). The Forgotten Revolution: the Priority Method: a restorative care model for older persons. *Trafford. Victoria, Canada*.

Markle-Reid M & Browne G. (2003). Conceptualisations of frailty in relation to older adults. *Journal of Advanced Nursing*. Vol 44:1 p 58-68.

Masters GN. (1982). A Rasch model for partial credit scoring. *Psychometrika*. Vol 47 p 147-174.

Maulden SA, Gassaway J, Horn D, Randell J, Smout MS & De Jongh G. (2005). Timing of initiation of rehabilitation after stroke. *Archives of Physical Medicine*. Vol 86:12 p 34–40.

Meagher D, O'Brian S, Pulella A & Brosnan P. (2009). Multidisciplinary activities in a community mental health service: relationship to Health of the Nation Outcomes Scale scores and diagnosis. *The Psychiatrist*. Vol 33 p 172-175.

- McCurren C. (2002). Assessment for depression among nursing home elders: Evaluation of the MDS mood assessment. *Geriatric Nursing*. Vol 23 p 103-108.
- McDaniel R & Bach C. (1996). Focus group research: the question of scientific rigor. *Rehabilitation Nursing Research*. Vol 5:2 p 53–59.
- Mc Dowell I & Newell C. (1996). *Measuring health: A guide to rating scales and questionnaires* (2nd ed.) New York: Oxford University Press.
- McGillis-Hall L, Doran D, Baker GR, Pink GH, Sidani S & O'Brien-Pallas L. (2001). The impact of nurse staffing mix models and organisational change strategies on patient, systems and nurse outcomes. Toronto, Canada: University of Toronto. Accessed 20 April 2009. [fhs.mcmaster.ca/nru/old%20nru/documents/web.McGillis%20Hall.pdf.]
- McLafferty I. (2004). Focus group interviews as a data collecting strategy. *Journal of Advance Nursing*. Vol 48:2 p 187-194.
- Morgan D. (1996). Focus groups. *Annual Review Sociology*. Vol 22 p 129–152. Annual review inc.
- Morgan DL. (1997). *Focus Groups as Qualitative Research*. SADE.
- Mosby's Medical Dictionary. (2009). *8th edition* ©, Elsevier.
- Mor V, Berg K, Angelelli J, Gilford D, Morris J & Moore T. (2003). The Quality of quality measurements in US nursing homes. *The Gerontologist (Special No. 2)*. Vol 43 p 34-46.
- Mukamel D & Spencer W. (2003). Quality Report Cards and Nursing Home Quality. *The Gerontologist (Special No. 2)*. Vol 43 p 58-66.
- Nelson A, Faan RN, Powell-Cope G, Palacios P, Luther SL, Black T, Hillman T, Christiansen B, Nathensen P, Coleman J. (2007) Nurse Staffing and Patient Outcomes in Inpatient Rehabilitation Settings. *Rehabilitation Nursing*. Vol 32:5 p 179–202.
- Nightingale F. (1859). *Notes on Nursing: What It Is and What It Is Not*. Harrison: London.

Nightingale F. (1863). *Notes on Hospitals* (3rd ed). London: Longmans, Green and Company.

Nightingale F. (1871). *Notes on Lying-In Institutions*, London: Longmans, Green and Company.

Nilsson AL, Sunnerhagen KS & Grimby G. 2005. Scoring alternatives for FIM in neurological disorders applying Rasch analysis. *Acta Neurological Scandinavica*. Vol 111 p 264–273.

Nouri FM & Lincon NB. (1987). An extended activities of daily living scale for stroke patients. *Clinical Rehabilitation*. Vol 1 p 301-305.

Omnibus Budget Reconciliation Act of 1987. Pub. L. No. 100-203. ~2.101 Stat. 1330. (1987) Accessed 10 October 2011. [<http://emedicine.medscape.com/article/317865>]

Open University. (1995). *The Passionate Statistician [video recording]*, Milton Keynes: Open University Educational Enterprises.

Orlando IJ. (1961). *The dynamic nurse patient relationship, function, process and principles*. New York. GP Putnam.

Orlando IJ & Dugan AB. (1989). Independent and dependent path: The fundamental issue for the nursing profession. *Nursing and Healthcare*. Vol 10:2 p 77-80.

Osborne JW. (2008). *Best practices in quantitative methods*. Sage Publishing. Thousand Oaks. CA. p 50-70.

Peile E. (2004). *Reflections from medical practice: balancing evidence-based practice with practice based evidence*. In, G. Thomas and R. Pring (Eds.) *Evidence-based Practice in Education*. Open University Press.

Pel-Littel R, Schuurmans MJ, Emmelot-Vonk MH & Verhaar HJJ. (2009). Frailty: defining and measuring of a concept. *The Journal of Nutrition Health and Aging*. Vol 13:4 p 39-394.

Polit DF & Beck CT. (2004). *Nursing Research. Principles and Methods*. Lippincott Williams & Wilkins, Philadelphia, PA.

Purkis J, Burgess P, Coombs T, Clarke A, Jones- Ellis D & Dickenson R. (2005). Routine measurements of outcomes in Australia's public sector mental health services. *Australia and New Zealand Health Policy*. Vol 2:8 [doi:1186/1743-8462-2-8]

Rantz MJ, Petroski GF, Madsen RW, Mehr DR, Popejoy L, Hicks IL, *et al.* (2000). *Setting thresholds for quality indicators derived from the MDS data for nursing homes quality improvement reports: An update*. Joint Commission Journal on Quality Improvement. Vol 26 p 101-110.

Rasch G. (1960). Probabilistic models for some intelligence and attainment tests. *Copenhagen: Denmark's Paedagogiske Institut*.

Resnick B, Cayo J & Pretzer-Aboff I. (2009). Implementation of the 6-Week Educational Component in the Res-Care Intervention. Process and Outcome. *Journal of Continuous Education in Nursing*. Vol 40:8 p 353–350.

Reuben DB, Laliberte L, Hiris J, Mor V. (1990). A hierarchical exercise scale to measure at the Advanced Activities of Daily Living (AADL) level. *Journal of the American Geriatric Society*. Vol 38 p 855-856.

RIC-FAS. Rehabilitation Institute of Chicago. (1987). *Rehabilitation Institute of Chicago Functional Assessment Scale*. Chicago Illinois.

Richards J. (2002). Foreword in Turner A, Foster M, Johnson SE (eds): *Occupational Therapy and Physical Dysfunction: Principles, Skills and Practice* (5th edn.) London. Churchill Livingstone.

Salvi G, Leese M & Slade M. (2005). Routine use of mental health outcome assessments: choosing the measure. *The British Journal of Psychiatry*. Vol 186 p 146-152.

Schnelle JF, Wood S, Schnelle ER, Simmonds SF. (2001). Measurement sensitivity and the minimal data set depression quality indicator. *The Gerontologist*. Vol 41p 401-405.

- Shin J & Scherer Y. (2009). Advantages and Disadvantages of using MDS Data in Nursing Research. *Journal of Gerontological Nursing*. Vol. 35:1 p 7-16.
- Simmonds M. (2002). Physical Function in Patients with Cancer: Psychometric Characteristics and Clinical Usefulness of a Physical Performance Test Battery. *Journal of Pain Symptom Management*. Vol 24:4 p 404-414.
- Sirios MJ, Lavoie A, Clermont E & Dionne E. (2004). Impact of transfer delays to rehabilitation in patients with severe trauma. *Archives of Physical Medicine*. Vol 65:2 p 184-191.
- Sijtjtsma K & Molenaar IW. (2002). *Introduction to non-parametric item response theory*. Thousand Oaks, California: Sage.
- Small H. (1999). "Florence Nightingale's statistical diagrams," paper presented at the Stats & Lamps Research Conference organised by the Florence Nightingale Museum at St. Thomas' Hospital, 18th March 1998. www.florence-nightingale.co.uk/
- Smith AJ. (2007). Embracing intuition in nursing practice. *Alabama Nurse*. Vol 34:3 p16-17.
- Smith B & MacVicar M. (1999). *Management of mobility and altered physical activity*. (In: Hinshaw AS, Feetham SL & Shaver JLF. (eds.) 1999. *Handbook of clinical nursing research*. Thousand Oaks, Ca.: Sage Publications. p.364).
- Smith EV Jr, & Andrich D. (2005). *Why Rasch Measurement or why don't (some) psychometricists get along?* Paper at the April Annual General Meeting of the American Educational Research Association. Montreal. Canada.
- Smith EV Jr. (2004). Evidence of reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In EV Smith Jr & RM Smith (Eds), *Introduction to Rasch Measurement*. Maple Grove, MN:JAM Press p 93-122.
- Smith EV Jr. (2005). Effect of item redundancy on Rasch item and person estimates. *Journal of Applied Measures*. Vol 6 p 147-163.

Smith RM, Linacre JM & Smith EV. (2011). Guidelines for manuscripts. Updates available on [<http://www.jampress.org>] (accessed 10 January 2011)

Stant AD, Buskens E, Jenner JA, Wiersma D & TenVergert E. (2007). Cost-effective analysis in severe mental illness: Outcomes measures selection. *Journal of Mental Health Policy and Economics*. Vol 10:2 p101-108.

Stevens SS. (1946). On the theory of scales and measurements. *Science*. Vol 103 p 677- 680.

Stineman MG, Tassoni CJ, Escarce JJ, Goin JE, Granger CV, Fiedler RC, &Williams SV. (1997). Development of function-related groups version 2.0: a classification system for medical rehabilitation. *Health Services Research*. Vol 32:4. p 529-548.

Stone MH. (2004). Substantive scale construction. In EV Smith Jr & RM Smith (Eds), *Introduction to Rasch Measurement*. Maple Grove, MN:JAM Press. (p. 201 - 225)

Strauss S E, Richardson W S, Glasziou P & Haynes R B. (2005). *Evidence-based medicine : How to practice and teach EBM (3rd ed.)*. Edinburgh ; New York: Elsevier/Churchill Livingstone.

Strawbridge WJ, Shema SJ, Balfour JL, Higby HR & Kaplan GA. (1998). Antecedents of frailty over three decades in an older cohort. *Journal of Gerontology Series B Psychological Sciences and Social Sciences*. Vol 53:1 p S9-16.

Tennant A & Conaghan P. (2007). The Rasch Measurement Model in Rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper. *Arthritis and rheumatism*. Vol 57:8 p 1358-1362.

Toomey M, Nicholson D & Carswell A. (1995). The clinical utility of the Canadian Occupational Performance Measure. *Canadian Journal of Occupational Therapy*. Vol 62:5 p 184-190)

- Trauer T, Callaly T & Herman H. (2009). Attitudes of mental health staff to routine outcomes measurement. *Journal of Mental Health*. Vol 8:4 p 288-296.
- Twinn S. (1998). An analysis of the effectiveness of focus groups as a method of qualitative data collection with Chinese populations in nursing research. *Journal of Advanced Nursing*. Vol 28:3 p 654–661.
- Unsworth C. (2000). Measuring the outcome of occupational therapy: tools and researches. *Australian Occupational Therapy Journal*. Vol 47:4 p 147-158.
- Van der Merwe JV. (2009). Clinical Governance - Clinical risk management workshop. Accessed 10 April 2011.
http://www.mediscor.net/Conferences/CRM0806/Clinical%20Governance_Prof%20vd%20Merwe.pdf.
- Van HS, van Geel AC, Geusens PP, Kessels A, Niewenhuijzen-Kreuseman AC & Brink PR. (2008). Bone and fall-related fracture risks in woman and men with recent clinical fracture. *Journal of Bone and Joint Surgery. American Volume*. Vol 90: 2 p 262-266.
- Verhalst ND & Glass CA. (1995). One parameter logistic model. In GH Fischer and IW Molenaar (Eds.). *Rasch models: Foundations, recent developments and applications*. p 215-237 New York : Springer.
- Victorian Department of Human Services. (2001). Sub-Acute / Acute Interface Project: *Final Report*. Victorian Government, Australia.
- Wade DT, Legh-Smith J, Hewer RL. (1985). Social activities after stroke: measurement and natural history using the Frenchey Activities Index. *International Rehabilitation Medicine*. Vol 7 p 176 –181.
- Weaver FM, Guihan M, Hynes DM, Byck G, Conrad KJ & Demakis JG. (1998). Prevalence of subacute patients in acute care: results of a study of VA hospitals. *Journal of Medical Systems*. Vol 22 p 161–172.
- Webb B. (2002). Using focus groups as a research method: a personal experience. *Journal of Nursing Management*. Vol 10 p 27–35.

WHO. (1980). *International classification of impairments, disabilities, and handicaps; a manual of classification relating to the consequences of disease*. World Health Organisation. Geneva.

Wiener J, Freiman M & Brown D. (2007). *Nursing Home Care Quality: Twenty Years after the Omnibus Budget Reconciliation Act of 1987. Kaiser Family Foundation Report*.

Wilson M. (2002). Book review. *Applied Psychological Measurement*. Vol 26:2.

Wind H, Gouttebargge V, Kuijer P, Sluiter J & Frings-Dresen M. (2006). The Utility of Functional Capacity Evaluation: the opinion of physicians and other experts in the field of return to work and disability claims. *International Archives of Occupational Environmental Health*. Vol 79 p 528-534.

WINSTEPS® version 3.70.1.1. (2010). www.winsteps.com

WINSTEPS®. (2010a). Processed 5 February 2010. SWREG Inc Order Number U795629501. [<https://usd.swreg.org/cgi-bin/r.cgi?o=795629501>].

WINSTEPS® (2011b). Accessed 10 January 2011. [<http://www.winsteps.com/winman/index.htm?history.htm>.]

Worrall J. (2007). Why there's no cause to randomize. *British Journal of the Philosophy of Science*. Vol 58 p 451-488.

Wright BD & Linacre JM. (1989). Observations are always ordinal; measures, however must be interval. *Archives of Physical Medicine and Rehabilitation*. Vol 70 p 857-860.

Wright BD & Masters GN. (1982). *Rating scale analysis*. Chicago. MESA Press.

Wright BD & Mok MM. (2004). An overview of the family of Rasch measurement models (p1-24). In EV Smith Jr & RM Smith (Eds). *Introduction to Rasch measurement*. Maple Grove, JAM Press.

Wright BD & Stone MH. (1979). *Best test design*. Chicago: MESA Press.

Wright BD. (1977). Misunderstanding the Rasch model. *Journal of Educational Measurement*. Vol 14:2 p 97-116.

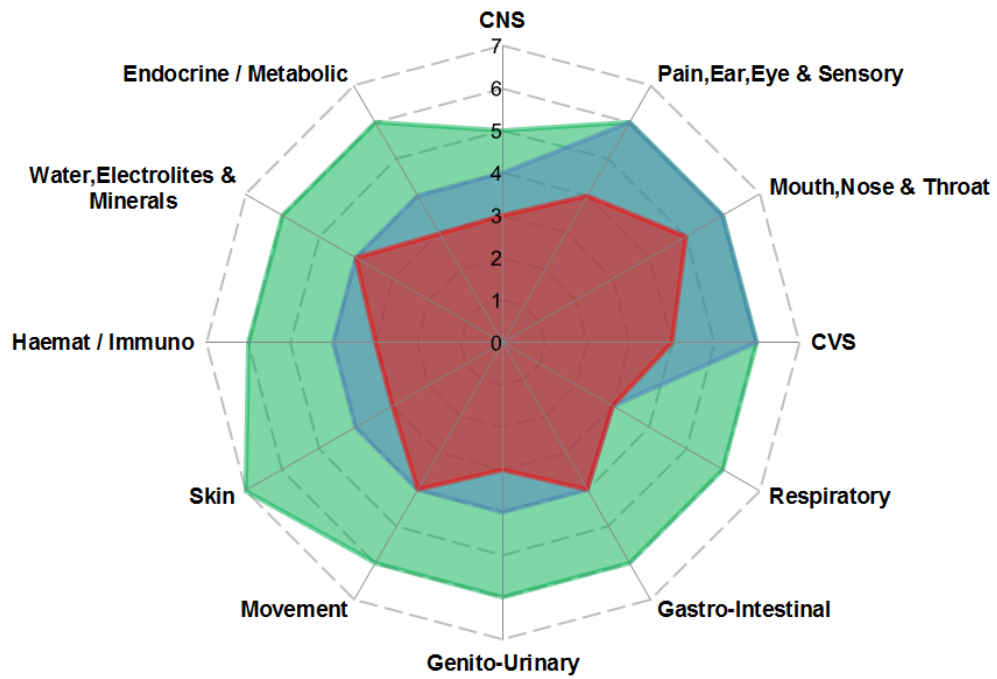
Wright BD. (1997). Fundamental measurement for outcomes evaluation. *Physical medicine and rehabilitation: State of the Art Review*. Vol 11:2 p 261-288.

Wright BD. (1997). Fundamental Measurement. *Rasch Measurement Transactions*. Vol 11:2 p 558.

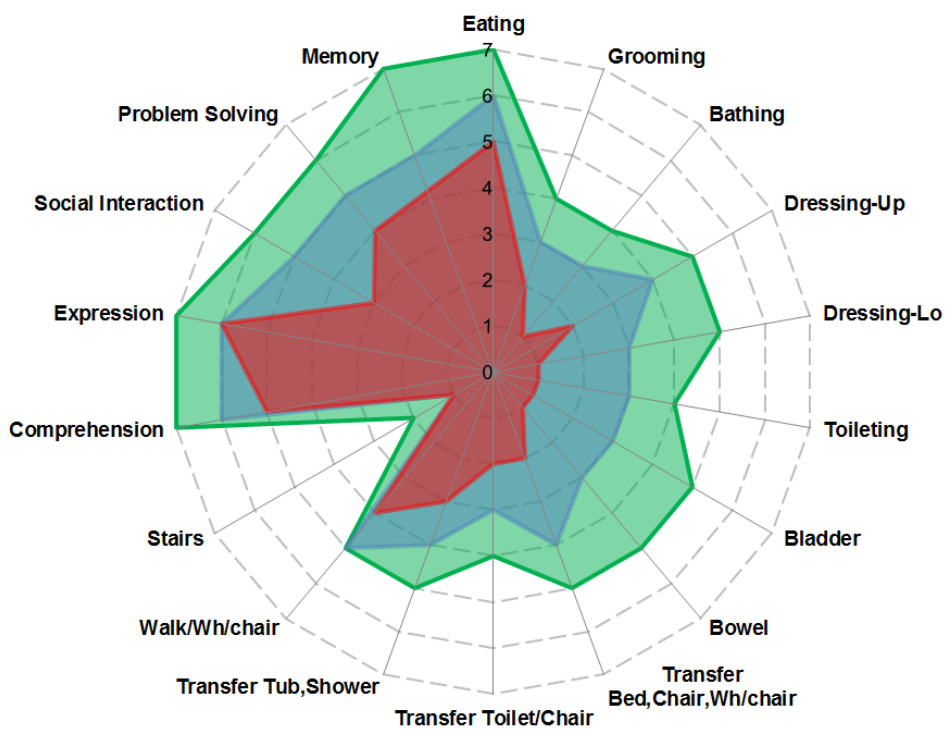
Yan Z. (2009). The development of a Rasch measurement physical fitness scale for Hong Kong primary school-aged students. PhD Thesis. James Cook University. Accessed 10 August 2011. [<http://eprints.jcu.edu.au/55569>]

Zimmerman DR. (2003). Using Quality indicators to improve residential and nursing home care. *International Journal of Geriatric Psychiatry*. Vol 18:3 p. 250–257.

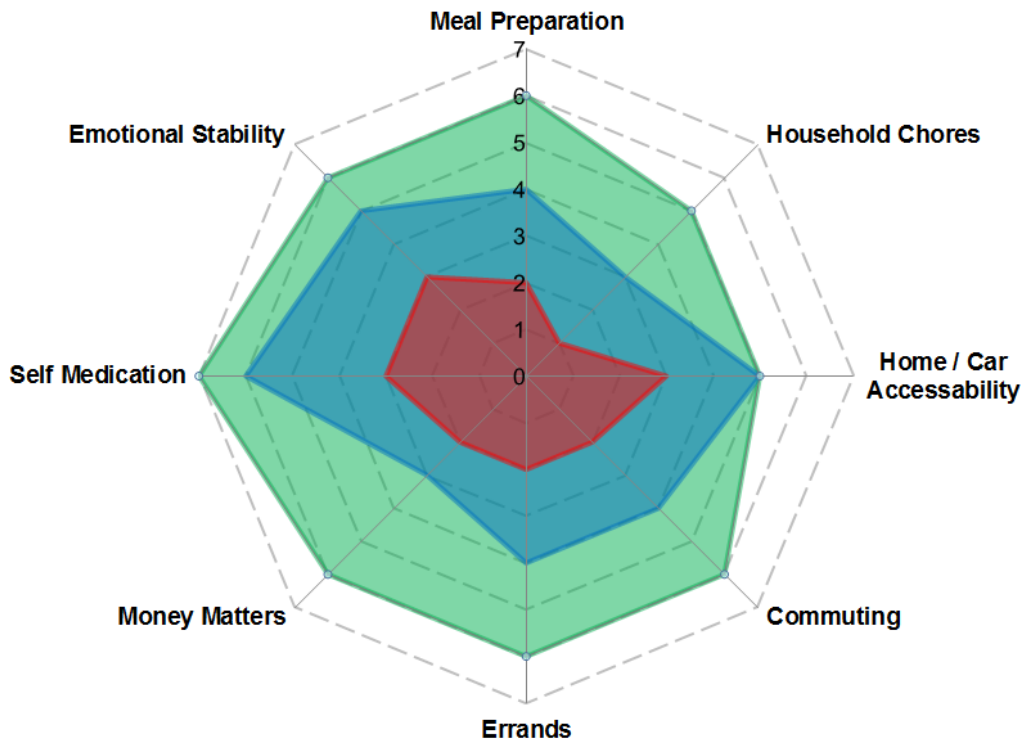
ALPHA Nursing Measure



BETA Nursing Measure



GAMMA Nursing Measure



DELTA Nursing Measure

